PCT/IL 2004/000335

REC'D 02 JUN 2004
WIPO PCT

מדינת ישראל
STATE OF ISRAEL

Ministry of Justice
Patent Office

משרד המשפטים
לשכת הפטנטים

This is to certify that annexed hereto is a true copy of the documents as originally deposited with the patent application particulars of which are specified on the first page of the annex.

זאת לתעודה כי רצופים בזה העתקים נכונים של המסמכים שהופקדו לכתחילה עם הבקשה לפטנט לפי הפרטים הרשומים בעמוד הראשון של הנספח.

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

LIBRARY ISRAEL PATENT OFFICE

This 29 APR 2004 היום

ממונה על הבוחנים
השם הפטנטים

Commissioner of Patents

נתאשר
Certified

לשימוש הלשכה
For Office Use

חוק הפטנטים, התשכ"ז -1967
PATENTS LAW, 5727-1967

מספר: 156273 Number

תאריך: 02-06-2003 Date
הוקדם/נדחה
Ante/Post-dated

בקשה לפטנט
Application for Patent

אני, (שם המבקש, מענו – ולגבי גוף מאוגד – מקום התאגדותו)
I (Name and address of applicant, and, in case of a body corporate, place of incorporation)

Yeda Research and Development Co. Ltd.
at the Weizmann Institute of Science
P.O. Box 95
76100 Rehovot

ידע חברה למחקר ופתוח בע"מ
ליד מכון ויצמן
ת.ד. 95
רחובות 76100

Inventors: Anthony FUTTERMAN; Joel L. SUSSMAN; Israel SILMAN; Michal HAREL; Hay DVIR; Lilly TOKER

הממציאים: אנטוני **פוטרמן**; יואל **זוסמן**; ישראל **סילמן**; מיכל **הראל**; חי **דביר**; לילי **טוקר**

בעל אמצאה מכח העברה ששמה הוא:
Owner, by virtue of **an Assignment from the inventors** of an invention, the title of which is:

(בעברית) **(Hebrew)**
שיטות לזיהוי תרופות למחלת גושר

(באנגלית) **(English)**
METHODS OF IDENTIFYING GAUCHER DISEASE DRUGS

hereby apply for a patent to be granted to me in respect thereof. מבקש בזאת כי ינתן לי עליה פטנט.

*בקשת חלוקה- Application for Division	*בקשת פטנט מוסף- Application for Patent of Addition	*דרישת דין קדימה Priority Claim		
מבקשת פטנט from Application	*לבקשה/לפטנט for Patent/Appl.	מספר/סימן Number/Mark	תאריך Date	מדינת האיגוד Convention Country
No. מס'	No. מס'	60/463,049	16 April 2003	USA
dated מיום	dated מיום			

*יפוי כח: ~~כללי~~/ מיוחד ~~הצהרה בזה~~/ עוד יוגש
P.O.A.: ~~general~~ / specific – ~~attached~~/ to be filed later
Has been filed in case הוגש בענין

Address for Service in Israel המען למסירת הודעות ומסמכים בישראל

G.E. Ehrlich (1995) Ltd.
28 Bezalel Street
52 521 Ramat Gan

ג'י. אי. ארליך (1995) בע"מ
רחוב בצלאל 28
רמת גן 52 521

חתימת המבקש
Signature of Applicant

היום 2 בחודש יוני שנת 2003
This 2nd of June 2003

For the Applicant
Attorney Docket No.: 26058

לשימוש הלשכה
For Office Use

טופס זה, כשהוא מוטבע בחותם לשכת הפטנטים ומרושם במספר ובתאריך ההגשה, הינו אישור להגשת הבקשה שפרטיה רשומים לעיל.
This form, impressed with the Seal of the Patent Office and indicating the number and date of filing, certifies the filing of the application. the particulars of which are set out above.

שיטות לזיהוי תרופות למחלת גושר

METHODS OF IDENTIFYING GAUCHER DISEASE DRUGS

METHODS OF IDENTIFYING GAUCHER DISEASE DRUGS

FIELD AND BACKGROUND OF THE INVENTION

The present invention relates to methods of identifying compounds capable of regulating the activity of mutant enzymes, to computing platforms capable of generating models representing 3D structures of crystallized enzymes, to crystallized enzymes, and to methods of crystallizing enzymes. In particular, embodiments of the present invention relate to methods of identifying compounds capable of correcting impaired enzymatic activity of mutant human glucocerebrosidase molecules associated with Gaucher disease, to computing platforms capable of generating models representing essentially complete experimentally determined 3D structures of crystallized human glucocerebrosidase polypeptide, to crystallized human glucocerebrosidase polypeptide, and to methods of crystallizing human glucocerebrosidase polypeptide.

Gaucher disease, the most common lysosomal storage disease (Meikle PJ. *et al.*, 1999. JAMA 281:249-54), is a highly debilitating disease occuring with a frequency of 1 in 40,000-60,000 in the general population, and 1 in 500-1,000 amongst Ashkenazi Jews [Beutler E. and Grabowski GA., in: "The Metabolic and Molecular Bases of Inherited Disease", Scriver CR. *et al.* (eds.), McGraw-Hill Inc., pp. 3635-3668 (2001)]. In Gaucher disease, deficiency in enzymatic activity of the lysosomal enzyme glucocerebrosidase due mutations in the enzyme [Beutler E. and Grabowski GA., in: "The Metabolic and Molecular Bases of Inherited Disease", Scriver CR. *et al.* (eds.), McGraw-Hill Inc., pp. 3635-3668 (2001)] leads to pathological lysosomal accumulation of glucosylceramide at various anatomic sites throughout the body, including the liver, spleen and bone marrow. The disease occurs in various forms, in particular Type 1 which is predominantly characterized by hepatosplenomegaly; and Types 2 and 3 which are characterized by early or chronic onset of severe neurological symptoms. Due to its genetic component, Gaucher disease represents a genetic testing dilemma for potential carriers.

Glucocerebrosidase (EC 3.2.1.45, acid beta-glucosidase, D-glucosyl-N-acylsphingosine glucohydrolase, glucosylceramidase) is a peripheral membrane protein which hydrolyzes the beta-glucosyl linkage of glucosylceramide in lysosomes, thereby generating beta-glucose and ceramide (Figure 1a). This enzymatic activity

requires the coordinate action of saposin C and negatively-charged lipids for maximal activity [Beutler E. and Grabowski GA., in: "The Metabolic and Molecular Bases of Inherited Disease", Scriver CR. *et al.* (eds.), McGraw-Hill Inc., pp. 3635-3668 (2001); Grabowski GA. *et al.*, 1990. Critical Rev Biochem Mol Biol. 25:385-414]. Based on sequence similarity, glucocerebrosidase is classified as a member of glycoside hydrolase family 30 (http://afmb.cnrs-mrs.fr/CAZY/GH_30.html).

Of the approximately 200 known glucocerebrosidase mutations, homozygosity for the common mutations N370S and L444P is associated with non-neuronopathic (Charrow J. *et al.*, 2000. Arch Intern Med. 160:2835-43) and neuronopathic [Erikson A. *et al.* in: "Gaucher's Disease", Zimran A. (ed.), Bailliere Tindall, London, pp. 711-723 (1997)] Gaucher disease, respectively. Mutation N370S is the most common mutation, accounting for about 70 and 25 percent of the mutant alleles in Ashkenazi Jewish and non-Jewish patients, respectively [Beutler E. and Grabowski GA., in: "The Metabolic and Molecular Bases of Inherited Disease", Scriver CR. *et al.* (eds.), McGraw-Hill Inc., pp. 3635-3668 (2001)]. Many of the glucocerebrosidase mutations (Figure 1d) are rare and restricted to a few individuals, and most partially or entirely decrease catalytic activity (Meivar-Levy, I. *et al.*, 1994. Biochem. J. 303:377-382) or may reduce glucocerebrosidase stability (Grace ME. *et al.*, 1994. J Biol Chem. 269:2283-2291).

Whole-enzyme replacement therapy with Cerezyme®, a recombinant variant of human glucocerebrosidase (Grabowski GA. *et al.*, 1995. Ann Intern Med. 122:33-9) is the main treatment for Type 1 Gaucher disease. Such treatment, however, is not curative, nor does it satisfactorily alleviate the symptoms of the disease. Furthermore, such whole-enzyme replacement therapy has numerous significant disadvantages, including: (i) administration of a molecule having various suboptimal pharmacokinetic characteristics, including suboptimal tissue penetration as a result of its large size, and suboptimal plasma membrane permeability, and *in-vivo* half-life due to its polypeptidic composition; (ii) incapacity to correct endogenous glucocerebrosidase enzyme activity, and thereby incapacity to therapeutically confer such activity with optimal spatial (cell type/subcellular location), temporal, and activity level regulation; (iii) for optimal therapeutic results, the need to administer the enzyme via injection, a painful, inconvenient, and expensive process; and (iv) elicitation of harmful immune responses against the administered enzyme in a

substantial proportion of treated subjects as a result of its polypeptidic/modified oligosaccharide chemical composition (http://www.gaucherregistry.com/safety/cerezyme_pi.html). Hence, there is a clearly felt need for novel/improved Gaucher disease drugs.

One approach for attempting to correct the defective enzymatic activity of the glucocerebrosidase Asn370Ser mutant which has been proposed in the prior art involves utilizing deoxynijirimycin (DNJ)-based compounds (Sawkar *et al.*, 2002. Proc Natl Acad Sci U S A. 99:15428-33). Such an approach, however, suffers from significant drawbacks, including: (i) the nonspecific inhibitory effect of DNJ compounds on enzymes that make and make glucosyl bonds, such as endoplasmic reticulum oligosaccharide-processing enzymes alpha-glucosidase I/II, ceramide glucosyl transferase, and both non-lysosomal and lysosomal glucocerebrosidase; (ii) the significant toxicity displayed by DNJ compounds; and (iii) the fact that DNJ compounds have not been demonstrated to have a significant restorative effect on the impaired enzymatic activity of any glucocerebrosidase mutant other than Asn370Ser, including a demonstrated ineffectiveness for correcting impaired enzymatic activity of the common Leu444Pro glucocerebrosidase mutant.

Thus, in sharp contrast to prior art Gaucher disease drugs, optimal Gaucher disease drugs, would be compounds having optimally small dimensions, a non-polypeptidic chemical composition, a capacity to correct impaired enzymatic activity of any of various glucocerebrosidase mutants, and would be capable of correcting the impaired enzymatic activity of glucocerebrosidase mutants associated with Gaucher disease *in-vivo* with optimal effectiveness and safety.

Ideally, such compounds could be computationally identified by obtaining sets of structure coordinates defining experimentally determined 3D structures of glucocerebrosidase at atomic resolution, using such sets of structure coordinates for producing computational platforms capable of generating models representing such 3D structures of human glucocerebrosidase at such atomic resolution, and using such computing platforms for computationally identifying compounds capable of interacting with mutant human glucocerebrosidase molecules associated with Gaucher disease in such a way as to correct impaired enzymatic activity thereof.

One prior art approach which has been employed involves using two-dimensional hydrophobic cluster analysis in attempts to provide sets of structure

coordinates defining predictive structures of human glucocerebrosidase molecules (Fabrega S. *et al.*, 2002. J Soc Biol. 196:151-60; Fabrega S. *et al.*, 2000. Glycobiology 10:1217-24).

Another approach which has been employed involves crystallizing Cerezyme® and analyzing such crystals via X-ray crystallography in attempts to generate X-ray diffraction data defining 3D structures of Cerezyme® suitable for computational identification of novel Gaucher disease drugs (Roeber D. *et al.*, 2003. Acta Cryst. D59:343-344).

These prior art approaches, however, have essentially failed. Approaches predictive methods based on two-dimensional hydrophobic cluster analysis are suboptimal due to the significant inaccuracies inherent to such predictive methods, and, in any case, have not provided sets of structure coordinates defining the structure of glucocerebrosidase, nor of significant portions thereof, at adequately high resolution, with satisfactory completeness, and with a satisfactory degree of accuracy. Furthermore, approaches involving crystallization of Cerezyme® have not succeeded in producing crystals capable of generating X-ray diffraction data defining structures of Cerezyme® or portions thereof.

Thus, all prior art approaches have failed to provide an adequate solution for producing computational platforms suitable for computationally identifying compounds capable of interacting with mutant human glucocerebrosidase molecules associated with Gaucher disease in such a way as to correct impaired enzymatic activity thereof.

There is thus a widely recognized need for, and it would be highly advantageous to have, a computing platform devoid of the above limitation.

SUMMARY OF THE INVENTION

According to one aspect of the present invention there is provided a method of crystallizing a glucocerebrosidase molecule, the method comprising: (a) partially deglycosylating the glucocerebrosidase molecule, thereby generating a partially glycosylated glucocerebrosidase molecule; and (b) subjecting the partially glycosylated glucocerebrosidase molecule to crystallization-inducing conditions, thereby crystallizing the glucocerebrosidase molecule.

According to further features in preferred embodiments of the invention

described below, the step of partially deglycosylating the glucocerebrosidase molecule is effected by treating the glucocerebrosidase molecule with N-glycosidase F.

According to still further features in the described preferred embodiments, an amino acid sequence of the glucocerebrosidase molecule comprises a first N-linked glycosylation consensus sequence, wherein the first N-linked glycosylation consensus sequence is attached to a sugar moiety comprising a monosaccharide or a disaccharide directly attached to the first N-linked glycosylation consensus sequence, and whereas the step of partially deglycosylating the glucocerebrosidase molecule is effected so as to leave the monosaccharide or the disaccharide attached to the first N-linked glycosylation consensus sequence.

According to still further features in the described preferred embodiments, the step of partially deglycosylating the glucocerebrosidase molecule is further effected so as to effectively fully deglycosylate all glycosylated N-linked glycosylation consensus sequences of the amino acid sequence of the glucocerebrosidase molecule other than the first N-linked glycosylation consensus sequence of the amino acid sequence of the glucocerebrosidase molecule.

According to still further features in the described preferred embodiments, the crystallization-inducing conditions comprise inducing evaporation of a crystallization solution containing the at least partially deglycosylated glucocerebrosidase molecule at a concentration of about 5 mg/ml, and a component selected from the group consisting of a buffer, a sodium salt, an ammonium salt, a sulfate salt, a chaotropic compound, a potassium salt, and a chloride ion.

According to still further features in the described preferred embodiments, the buffer is a Zwitterionic buffer or an acetate buffer.

According to still further features in the described preferred embodiments, the buffer is 2-morpholinoethanesulfonic acid buffer or sodium acetate buffer.

According to still further features in the described preferred embodiments, the crystallization solution contains the buffer at a concentration of about 0.5 millimolar or about 0.05 molar.

According to still further features in the described preferred embodiments, the solution of a buffer has a pH of about 6.6 or about 4.6.

According to still further features in the described preferred embodiments, the

sodium salt is sodium chloride.

According to still further features in the described preferred embodiments, the crystallization solution contains the sodium salt at a concentration of about 0.05 molar.

According to still further features in the described preferred embodiments, the ammonium salt is ammonium sulfate.

According to still further features in the described preferred embodiments, the crystallization solution contains the ammonium salt at a concentration of about 0.5 molar.

According to still further features in the described preferred embodiments, the crystallization solution contains the sulfate salt at a concentration of about 0.5 molar.

According to still further features in the described preferred embodiments, the chaotropic compound is guanidine hydrochloride.

According to still further features in the described preferred embodiments, the crystallization solution contains the chaotropic compound at a concentration of about 0.085 molar.

According to still further features in the described preferred embodiments, the potassium salt is potassium chloride.

According to still further features in the described preferred embodiments, the crystallization solution contains the potassium salt at a concentration of about 0.01 molar.

According to still further features in the described preferred embodiments, the crystallization solution contains the chloride ion at a concentration of about 0.06 molar.

According to still further features in the described preferred embodiments, the crystallization solution has a pH of about 4.6.

According to still further features in the described preferred embodiments, inducing evaporation of the crystallization solution is effected at a temperature of about 22 degrees centigrade.

According to another aspect of the present invention there is provided a method of identifying a compound capable of correcting an impaired enzymatic activity of a mutant glucocerebrosidase molecule, the method comprising: (a) obtaining a first set of structure coordinates, the first set of structure coordinates

defining a 3D structure of a glucocerebrosidase molecule capable of displaying normal enzymatic activity or a portion thereofset of structure coordinates defines ; (b) computationally generating using the first set of structure coordinates a second set of structure coordinates, the second set of structure coordinates defining a predicted 3D structure of the mutant glucocerebrosidase molecule or a portion thereof; and (c) computationally identifying, using the second set of structure coordinates, a compound capable of interacting with the mutant glucocerebrosidase molecule in such a way as to correct the impaired enzymatic activity thereof, thereby identifying the compound capable of correcting the impaired enzymatic activity of the mutant glucocerebrosidase molecule.

According to further features in preferred embodiments of the invention described below, the step of computationally identifying using the second set of structure coordinates a compound capable of interacting with the mutant glucocerebrosidase molecule is effected further using the first set of structure coordinates.

According to still further features in the described preferred embodiments, the method of identifying a compound capable of correcting an impaired enzymatic activity of a mutant glucocerebrosidase molecule further comprises biochemically qualifying a capacity of the compound to correct the impaired enzymatic activity of the mutant glucocerebrosidase molecule.

According to still further features in the described preferred embodiments, the amino acid sequence of the glucocerebrosidase molecule capable of displaying normal enzymatic activity is set forth in SEQ ID NO: 1.

According to still further features in the described preferred embodiments, the amino acid sequence of the glucocerebrosidase molecule capable of displaying normal enzymatic activity is composed of 497 amino acid residues, and the portion of the glucocerebrosidase molecule capable of displaying normal enzymatic activity comprises a set of amino acid residues of the amino acid sequence of the glucocerebrosidase molecule having normal activity having amino acid sequence coordinates selected from the group consisting of: (i) 76, 81, 285, 312, 314, 320, 324, 325, 336, 364–378, 423, and 433; (ii) 244–247, and 390–397; (iii) 20, 21, 95–100, and 404–411; (iv) 65–67, 440–447, 460–464, 468, and 469; (v) 360–366, 443–446, 460–467, and 484–89; and (vi) 33–35, 69, 71, 450–456, 474–478, and 493–497.

According to still further features in the described preferred embodiments, the amino acid sequence of the mutant glucocerebrosidase molecule is composed of 497 amino acid residues, and the portion of the mutant glucocerebrosidase molecule comprises a set of amino acid residues of the amino acid sequence of the mutant glucocerebrosidase molecule having amino acid sequence coordinates selected from the group consisting of: (i) 76, 81, 285, 312, 314, 320, 324, 325, 336, 364–378, 423, and 433; (ii) 244–247, and 390–397; (iii) 20, 21, 95–100, and 404–411; (iv) 65–67, 440–447, 460–464, 468, and 469; (v) 360–366, 443–446, 460–467, and 484–89; and (vi) 33–35, 69, 71, 450–456, 474–478, and 493–497.

According to still further features in the described preferred embodiments, the first set of structure coordinates comprises a set of structure coordinates set forth in Table 4, 5, 6, 7, 8, 9, and/or 10.

According to still further features in the described preferred embodiments, the amino acid sequence of the mutant glucocerebrosidase molecule is set forth in SEQ ID NO: 2, 3, 4, 5, 6, or 7.

According to still further features in the described preferred embodiments, the second set of structure coordinates comprises a set of structure coordinates set forth in Table 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, and/or 22.

According to still further features in the described preferred embodiments, the glucocerebrosidase molecule capable of displaying normal enzymatic activity is a crystallized glucocerebrosidase molecule.

According to yet another aspect of the present invention there is provided a computing platform capable of generating a model representing a 3D structure of a glucocerebrosidase molecule or a portion thereof, the computing platform comprising: (a) a data-storage device storing data comprising a set of structure coordinates defining the 3D structure of the glucocerebrosidase molecule or the portion thereof; and (b) a processing unit being for generating the model representing the 3D structure from the data stored in the data-storage device.

According to further features in preferred embodiments of the invention described below, the glucocerebrosidase molecule is a glucocerebrosidase molecule capable of displaying normal enzymatic activity, or is a mutant glucocerebrosidase molecule.

According to still further features in the described preferred embodiments, the

amino acid sequence of the glucocerebrosidase molecule is set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, or 7.

According to still further features in the described preferred embodiments, the amino acid sequence of the glucocerebrosidase molecule is composed of 497 amino acid residues, and the portion of the glucocerebrosidase molecule comprises a set of amino acid residues of the amino acid sequence having amino acid sequence coordinates selected from the group consisting of: (i) 76, 81, 285, 312, 314, 320, 324, 325, 336, 364–378, 423, and 433; (ii) 244–247, and 390–397; (iii) 20, 21, 95–100, and 404–411; (iv) 65–67, 440–447, 460–464, 468, and 469; (v) 360–366, 443–446, 460–467, and 484–89; and (vi) 33–35, 69, 71, 450–456, 474–478, and 493–497.

According to still further features in the described preferred embodiments, the set of structure coordinates comprises a set of structure coordinates set forth in Table 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, and/or 22.

According to still further features in the described preferred embodiments, the glucocerebrosidase molecule is a crystallized glucocerebrosidase molecule.

According to still another aspect of the present invention there is provided a composition-of-matter comprising a crystallized glucocerebrosidase molecule, wherein the crystallized glucocerebrosidase molecule is characterized by an X-ray diffraction capacity enabling generation of a set of structure coordinates defining a 3D structure of the glucocerebrosidase molecule or a portion thereof, to a resolution of 2.9 angstroms or higher; and/or wherein an amino acid sequence of the glucocerebrosidase molecule is partially glycosylated.

According to further features in preferred embodiments of the invention described below, the crystallized glucocerebrosidase molecule is characterized by unit cell dimensions of a = about 107.7 angstroms, b = about 285.2 angstroms and c = about 91.8 angstroms.

According to still further features in the described preferred embodiments, the crystallized glucocerebrosidase molecule is characterized by a crystal space group of $C222_1$.

According to a further aspect of the present invention there is provided a computer-readable medium comprising, in a retrievable format, data including a set of structure coordinates defining a 3D the set of structure coordinates defines the 3D structure at a resolution of 2.9 angstroms or higher, and/or wherein an amino acid

sequence of the glucocerebrosidase molecule is partially glycosylated.

According to yet a further aspect of the present invention there is provided a computer generated model representing a 3D structure of a glucocerebrosidase molecule or a portion thereof, wherein the model represents the glucocerebrosidase molecule or a portion thereof at a resolution of 2.9 angstroms or higher, and/or wherein the an amino acid sequence of the glucocerebrosidase molecule is partially glycosylated.

According to still further features in the described preferred embodiments, the partially glycosylated amino acid sequence comprises a first N-linked glycosylation consensus sequence attached to a sugar.

According to still further features in the described preferred embodiments, the set of structure coordinates defines the 3D structure to a resolution of 2.9 angstroms or higher

According to still further features in the described preferred embodiments, the amino acid sequence of the glucocerebrosidase molecule is partially glycosylated.

According to still further features in the described preferred embodiments, the glucocerebrosidase molecule is capable of displaying normal enzymatic activity.

According to still further features in the described preferred embodiments, the amino acid sequence of the glucocerebrosidase molecule is set forth in SEQ ID NO: 1.

According to still further features in the described preferred embodiments, the set of structure coordinates comprises a set of structure coordinates set forth in Table 4, 5, 6, 7, 8, 9, and/or 10.

The present invention successfully addresses the shortcomings of the presently known configurations by providing: (i) a composition-of-matter comprising crystallized human glucocerebrosidase having an X-ray diffraction capacity enabling generation of structure coordinates defining the 3D structure of human glucocerebrosidase with optimal resolution; (ii) a method of generating such a composition-of-matter; (iii) a computing platform utilizing such structure coordinates for generating, at optimally high resolution, models representing the experimentally determined, essentially complete 3D structure of human glucocerebrosidase, including that of mutable portions thereof, and the predicted, essentially complete, 3D structure of mutant human glucocerebrosidase molecules, including that of mutant portions thereof; and (iv) methods of utilizing such structure coordinates for computationally

identifying compounds capable of serving as optimal Gaucher disease drugs.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the patent specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is herein described, by way of example only, with reference to the accompanying drawings. With specific reference now to the drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of the preferred embodiments of the present invention only, and are presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of the invention. In this regard, no attempt is made to show structural details of the invention in more detail than is necessary for a fundamental understanding of the invention, the description taken with the drawings making apparent to those skilled in the art how the several forms of the invention may be embodied in practice.

In the drawings:

FIG. 1a is a prior art diagram depicting hydrolysis of the beta-glucosyl linkage of glucosylceramide by glucocerebrosidase in lysosomes to generate beta-glucose and ceramide.

FIG. 1b is a 3D structure diagram depicting the 2.0 angstrom atomic 3D structure of a human glucocerebrosidase molecule capable of displaying normal enzymatic activity. Domain I is shown in magenta and contains the 2 disulfide bridges, whose S-atoms are shown as green balls. The glycosylation site at N19 is depicted as balls and sticks. Domain II, an Ig-like domain, is shown in green. The catalytic domain (domain III), a TIM barrel, is shown in cyan, and the active site residues E235 and E340 shown as balls and sticks. The six most common

glucocerebrosidase mutations are shown as balls, with those that predispose to severe (i.e., Type 2 or 3) and mild (i.e., Type 1) disease in red and yellow, respectively.

FIG. 1c is a 2D topology diagram depicting the 2D topology map of the structure of a human glucocerebrosidase molecule capable of displaying normal enzymatic activity. The diagram is consistent with a 3D-view looking down the opening of the active site pocket, as in panel A. All connecting loops in the diagram are of arbitrary length. Alpha-helices and beta-strands of Domain III are numbered according to their position in the sequence. For clarity, sequence coordinates for certain key positions are shown in the connecting loops, and secondary elements four residues or smaller are not shown.

FIG. 1d is a sequence diagram depicting the primary amino acid sequence of the 497 amino acid residues of human glucocerebrosidase with normal enzymatic activity (SEQ ID NO: 8), showing the amino acid positions thereof associated with pathological effects. Positions of mutations associated with severe disease are shown in red, mild disease in yellow, and those for which clinical data documenting disease severity are lacking, in blue. Only single amino acid substitutions are included, with frameshifts and splices excluded as enzyme is not expressed in most of these cases. Helices are indicated by cylinders and beta-strands by arrows and colored according the domains as shown in Figure 1b, above.

FIG. 1e is a structure diagram depicting the distribution of reported single amino acid substitutions in the 3D structure of glucocerebrosidase that lead to Gaucher disease. Mutations reported to cause severe disease are shown in red, mild disease in yellow, and those for which clinical data documenting disease severity are lacking, in blue, according to the same color code employed in Figure 1d, above. In some cases, phenotype assignment as mild (type 1) or severe (types 2 and 3) is based on a few, or sometimes only one individual. The phenotype of a number of mutations is not known since the mutation was detected in genomic DNA and data concerning disease severity may not have been available. The active site glutamate residues are shown as black sticks.

FIG. 2a is a 3D structure diagram depicting the catalytic and glucone binding site of a human glucocerebrosidase molecule capable of displaying normal enzymatic activity. Amino acid residues near the catalytic glutamates (ball-and-stick models) are rendered as stick models. H-bonds are shown as dashed lines for those residues in

contact distance to the glutamates. These residues may be directly involved in catalysis or modulate the protonation states of the carboxyl groups. The others are near the docked glucosyl moiety (see Figure 2b, below), and may thus stabilize its interaction with glucocerebrosidase.

FIG. 2b is a diagram depicting the 3D surface of a human glucocerebrosidase molecule capable of displaying normal enzymatic activity with a model of the docked glucosylceramide substrate. Hydrophobic residues (W, F, Y, L, I, V, M and C; Hopp TP. and Woods KR., 1981. Proc Natl Acad Sci U S A. 78:3824-8) are shown in blue, and the active site residues (E235 and E340) in yellow. Glucosylceramide is shown in CPK format (carbon and oxygen atoms in green and red, respectively). The model of the docked substrate was based on the coordinates of galactosylceramide (Nyholm PG. *et al.*, 1990. Chem Phys Lipids 52:1-10) modified for glucosylceramide. The surface diagram of glucocerebrosidase was rendered using PYMOL software (htttp://www.pymol.org).

FIG. 3 is a 3D structure diagram depicting a cluster of mutations in helix 7 in human glucocerebrosidase. The transparent ribbon diagram shows the three domains of glucocerebrosidase as in Figure 1a, but rotated by about 90 degrees around the x-axis to look down helix 7 (colored red). The amino acids mutated in Gaucher disease on this helix (R359, Y363, S366, T369, N370) are depicted as red balls and sticks. The active site residues E235 and E340 are depicted with carbon atoms as yellow balls and oxygen atoms as red balls.

FIG. 4 is a schematic diagram depicting a computing platform for generating a model representing a 3D structure of a glucocerebrosidase molecule or a portion thereof.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is of a method of identifying a compound capable of correcting an impaired enzymatic activity of a mutant glucocerebrosidase molecule, a composition-of-matter comprising a crystallized glucocerebrosidase molecule having X-ray diffraction capacity enabling generation of structure coordinates defining a 3D structure of such a molecule or portion thereof, a method of generating such a crystallized molecule, a computing platform capable of generating a model representing a 3D structure of a glucocerebrosidase molecule or portion thereof, a

computer-readable medium comprising, in a retrievable format, data including a set of structure coordinates defining such a structure, and a computer generated model representing such a structure. Specifically, the present invention can be used for identifying a compound capable of optimally correcting an impaired enzyme activity of a mutant glucocerebrosidase molecule associated with Gaucher disease, such compound having optimal physical, chemical, and/or biological characteristics for use as a drug. As such, the present invention can be used for identifying a Gaucher disease drug being markedly superior to all prior art Gaucher disease drugs.

Before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details set forth in the following description or exemplified by the Examples. The invention is capable of other embodiments or of being practiced or carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein is for the purpose of description and should not be regarded as limiting.

Gaucher disease is a devastating disorder resulting from impaired glucocerebrosidase enzymatic activity, and for which the main treatment, whole-enzyme replacement therapy, has numerous and significant drawbacks, including suboptimal effectiveness, harmful side-effects, and a undesirable preferred route of administration, as described above.

One approach for restoring defective enzymatic activity of a glucocerebrosidase mutant has been attempted in the prior art using deoxynijirimycin (DNJ)-based chemicals. Such an approach, however, suffers from the significant drawbacks of employing a type of compound having broadly non-specific inhibitory effects on numerous vital enzymes, of not having demonstrated effectiveness in correcting defective activity in numerous common glucocerebrosidase mutants, and of being potentially highly toxic.

Ideally, optimal Gaucher disease drugs could be computationally identified using sets of structure coordinates defining the 3D atomic structure of glucocerebrosidase. Such sets of structure coordinates could be used for computationally generating a model representing the predicted 3D structure of a human glucocerebrosidase mutant associated with Gaucher disease. Such models could in turn be used for computationally identifying a compound capable of optimally correcting a defective enzymatic activity of the mutant glucocerebrosidase

molecule *in-vivo*.

Various approaches for obtaining sets of structure coordinates representing the 3D atomic structure of human glucocerebrosidase have been described by the prior art.

One such approach involves using two-dimensional hydrophobic cluster analysis in attempts to provide sets of structure coordinates defining predictive structures of human glucocerebrosidase molecules. Another approach involves crystallizing Cerezyme® and analyzing such crystals via X-ray crystallography in attempts to generate X-ray diffraction data defining 3D structures of Cerezyme®.

Such prior art approaches, however, have failed to provide a set of structure coordinates defining a 3D atomic structure of glucocerebrosidase or a portion thereof enabling computational identification of optimal Gaucher disease drugs.

While reducing the present invention to practice, the present inventors succeeded in crystallizing a human glucocerebrosidase molecule having normal enzymatic activity of which crystallographic analysis was used for successfully generating for the first time a set of structure coordinates defining the essentially complete structure of the crystallized glucocerebrosidase molecule at atomic resolution. The present inventors further used this set of structure coordinates to generate for the first time sets of structure coordinates defining with optimally high resolution and accuracy essentially complete predicted 3D structures of mutant glucocerebrosidase molecules associated with Gaucher disease, including mutant portions thereof. While conceiving the present invention, the present inventors hypothesized that such sets of structure coordinates defining predicted 3D structures of mutant glucocerebrosidase molecules, in particular mutant portions thereof, could be used for identifying a compound capable of correcting *in-vivo*, with optimal efficacy and safety, the impaired enzymatic activity of a mutant glucocerebrosidase molecule associated with Gaucher disease.

Thus, according to the present invention there is provided a composition-of-matter comprising a crystallized glucocerebrosidase molecule having an X-ray diffraction capacity enabling generation of a set of structure coordinates defining a 3D structure of the glucocerebrosidase molecule or a portion thereof.

The composition-of-matter of the present invention can be used for generating for the first time a set of structure coordinates defining at atomic resolution an

essentially complete 3D structure of a correctly folded glucocerebrosidase molecule, and in particular a portion thereof comprising an amino acid residue whose mutation may be associated with Gaucher disease (hereinafter referred to as "mutable portion"). Hence, as described hereinbelow, such a set of structure coordinates can be used for the first time for generating a computational model representing at atomic resolution a 3D structure of such a glucocerebrosidase molecule or a portion thereof. By using such a suitable computing platform, the data provided therein can be used for the first time for generating a set of structure coordinates defining, at atomic resolution and with optimal accuracy, an essentially complete predicted 3D structure of a glucocerebrosidase molecule having a mutated amino acid sequence associated with Gaucher disease (referred to hereinafter as "mutant glucocerebrosidase molecule"), or a portion thereof comprising a mutated amino acid residue associated with Gaucher disease (referred to hereinafter as "mutant portion"). The present inventors further hypothesized that such a computing platform could be used, preferably in conjunction with the set of structure coordinates defining the structure of the correctly folded glucocerebrosidase molecule, or a portion thereof, for computationally identifying a molecule capable of optimally correcting an impaired enzymatic activity of the mutant glucocerebrosidase molecule, and hence for identifying an optimal Gaucher disease drug.

Alternate art nomenclature for the chemical name "glucocerebrosidase" includes acid beta-glucosidase, D-glucosyl-N-acylsphingosine glucohydrolase, and glucosylceramidase. Gaucher disease may also be referred to in the art as glucosylceramide storage disease; GSDI)

As used herein, the phrase "3D structure" refers to the spatial positioning of a set of atoms, including a set of atoms which are not directly interconnected, and is used interchangeably with the phrase "three dimensional structure".

As used herein, an "essentially complete structure" of a glucocerebrosidase molecule or a portion thereof corresponds to a structure covering at least 95 percent of the complete structure of the glucocerebrosidase molecule or the portion thereof, respectively.

Preferably the crystallized glucocerebrosidase molecule of the present invention is essentially of mammalian origin, most preferably essentially of human origin.

The crystallized glucocerebrosidase molecule of the present invention may be characterized by:

(i) an X-ray diffraction capacity enabling generation of a set of structure coordinates defining a 3D structure of the crystallized glucocerebrosidase molecule or a portion thereof at any of various resolutions at least as high as 2.9 angstroms;

(ii) an X-ray diffraction capacity enabling generation of a set of structure coordinates defining an essentially complete 3D structure of the glucocerebrosidase molecule or essentially any portion thereof;

(iii) any of various unit cell dimensions;

(iv) any of various crystal space groups;

(v) a glucocerebrosidase molecule having any of various capacities to display enzymatic activity;

(vi) a glucocerebrosidase molecule having any of various amino acid sequences; and/or

(vii) a glucocerebrosidase molecule having an amino acid sequence characterized by any of various glycosylation patterns.

As described hereinabove, the crystallized glucocerebrosidase molecule may be characterized by an X-ray diffraction capacity enabling generation of a set of structure coordinates defining a 3D structure of the glucocerebrosidase molecule or a portion thereof at any of various resolutions at least as high as 2.9 angstroms.

The crystallized glucocerebrosidase molecule preferably has an X-ray diffraction capacity enabling generation of a set of structure coordinates defining a 3D structure of the glucocerebrosidase molecule or a portion thereof at a resolution of 2.9 angstroms or higher, more preferably at a resolution of 2.8 angstroms or higher, more preferably at a resolution of 2.8 angstroms or higher, more preferably at a resolution of 2.7 angstroms or higher, more preferably at a resolution of 2.6 angstroms or higher, more preferably at a resolution of 2.5 angstroms or higher, more preferably at a resolution of 2.4 angstroms or higher, more preferably at a resolution of 2.3 angstroms or higher, more preferably at a resolution of 2.2 angstroms or higher, more preferably at a resolution of 2.1 angstroms or higher, and most preferably at a resolution of 2.0 angstroms.

As is described and illustrated in Example 1 of the Examples section below, the crystallized glucocerebrosidase molecule of the present invention has an X-ray

diffraction capacity enabling for the first time generation of a set of structure coordinates defining an essentially complete 3D structure of the glucocerebrosidase molecule or essentially any portion thereof at a resolution of 2.0 angstroms. Such a set of structure coordinates of the present invention can be used for generating for the first time a model representing an essentially complete 3D atomic structure of a glucocerebrosidase molecule.

X-ray crystallography:

Utilizing an X-ray diffraction capacity of a crystallized molecule, such as that of an enzyme, for generating a set of structure coordinates defining a 3D structure of the crystallized molecule or a portion thereof is well within the purview of the ordinarily skilled artisan. The process of utilizing an X-ray diffraction capacity of a crystallized molecule, such as that of an enzyme, for generating a set of structure coordinates defining a 3D structure of the crystallized molecule or a portion thereof is referred to in the art as X-ray crystallography. X-ray crystallography is generally effected by exposing crystals to an X-ray beam and collecting the resultant X-ray diffraction data. This process usually involves the measurements of many tens of thousands of data points over a period of one to several days depending on the crystal form and the resolution of the data required. The crystals diffract the rays, creating a geometrically precise pattern of spots recorded on photographic film or electronic detectors. The distribution of atoms within the crystal influences the pattern of spots. The quality of protein crystals is determined by the ability of the crystal to scatter X-rays of wavelengths (typically 1.0-1.6 angstroms) suitable for determining the atomic coordinates of the macromolecule. The measure of the quality is determined as a function of the highest angle of scatter (the ultimate or intrinsic resolution) and according to Bragg's Law: $(n)(lambda) = (2d)(sin\theta)$, where θ is the angle of incidence of the reflected X-ray beam, d is the distance between atomic layers in a crystal, *lambda* is the wavelength of the incident X-ray beam, and n is an integer; d may be determined, and represents the resolution of the crystal form in angstroms. Thus, this measurement is routinely used to judge the ultimate usefulness of protein crystals. Group theory shows that there are 230 unique ways in which chemical substances, proteins or otherwise, may assemble in 3D to form crystals. These are called the 230 "space groups". The designation of the space group in addition to the unit cell constants (which define the explicit size and shape of the cell which repeats

periodically within the crystal) is routinely used to uniquely identify a crystalline substance. Certain conventions have been established to ensure the proper identification of crystalline materials and these conventions have been set forth and documented in the International Tables for Crystallography, incorporated herein by reference.

While X-ray crystallography of a crystallized glucocerebrosidase molecule of the present invention may be effected in various ways, it is preferably effected according to the protocol set forth in Example 1 of the Examples section.

As described hereinabove, the crystallized glucocerebrosidase molecule of the present invention may be characterized by any of various unit cell dimensions.

Preferably, the crystallized glucocerebrosidase molecule of the present invention is characterized by unit cell dimensions of about a = about 107.7 angstroms, b = about 285.2 angstroms, and c = about 91.8.

As used herein the term "about" refers to ± 10 percent.

As described hereinabove, the crystallized glucocerebrosidase molecule of the present invention may be characterized by any of various crystal space groups.

Preferably, the crystallized glucocerebrosidase molecule of the present invention is characterized by a crystal space group of $C222_1$.

As described hereinabove, the crystallized glucocerebrosidase molecule of the present invention may be capable of displaying any of various enzymatic activity levels.

Preferably, the crystallized glucocerebrosidase molecule of the present invention is capable of displaying normal enzymatic activity.

As used herein, the phrase "normal enzymatic activity" refers to a specific glucocerebrosidase enzymatic activity level normally displayed by a wild-type glucocerebrosidase molecule (SEQ ID NO: 8). As is widely understood in the art glucocerebrosidase enzymatic activity is hydrolysis of the beta-glucosyl linkage of glucosylceramide (for example, refer to Figure 1 of the Examples section for a schematic diagram depicting the enzymatic reaction) or of a synthetic beta-glucoside.

A specific glucocerebrosidase enzymatic activity level displayed by a wild-type glucocerebrosidase molecule will typically be:

(i) 2320-2560 nmol of the synthetic beta-glucoside 2-N-(NBD-dodecanoyl-)-sphingosyl-1-O-beta-glucoside hydrolyzed per hour per milligram of

glucocerebrosidase, as determined according to "Method 1" set forth in Grabowski *et al.*, 1986. 261:8263;

(ii) 2090-2178 nmol of the synthetic beta-glucoside 2-N-(NBD-dodecanoyl-)-sphingosyl-1-O-beta-glucoside hydrolyzed per hour per milligram of glucocerebrosidase, as determined according to "Method 2" set forth in Grabowski *et al.*, 1986. 261:8263.

(iii) 2,790-3,610 nmol of the synthetic beta-glucoside 2-N-(NBD-hexanoyl-)-sphingosyl-1-O-beta-glucoside hydrolyzed per hour per milligram of glucocerebrosidase, as determined according to "Method 2" set forth in Grabowski *et al.*, 1986. 261:8263;

(iv) 729-818 nmol of the synthetic beta-glucoside 4-methylumbelliferyl-11-O-beta-glucoside hydrolyzed per hour per milligram of glucocerebrosidase, as determined according to "Method 2" set forth in Grabowski *et al.*, 1986. 261:8263;

(v) 1,973-2,460 nmol of the synthetic beta-glucoside 4-methylumbelliferyl-9-O-beta-glucoside hydrolyzed per hour per milligram of glucocerebrosidase, as determined according to "Method 2" set forth in Grabowski *et al.*, 1986. 261:8263;

(vi) 1,820-2,124 nmol of the synthetic beta-glucoside 4-methylumbelliferyl-7-O-beta-glucoside hydrolyzed per hour per milligram of glucocerebrosidase, as determined according to "Method 2" set forth in Grabowski *et al.*, 1986. 261:8263;

(vii) 2,240-2,280 nmol of the synthetic beta-glucoside 4-methylumbelliferyl-1-O-beta-glucoside hydrolyzed per hour per milligram of glucocerebrosidase, as determined according to "Method 1" set forth in Grabowski *et al.*, 1986. 261:8263;

(viii) 2,038-2,430 nmol of the synthetic beta-glucoside 4-methylumbelliferyl-1-O-beta-glucoside hydrolyzed per hour per milligram of glucocerebrosidase, as determined according to "Method 2" set forth in Grabowski *et al.*, 1986. 261:8263; and/or

(ix) 830-1,000 nmol of the synthetic beta-glucoside 4-methylumbelliferyl-1-O-beta-glucoside hydrolyzed per hour per milligram of glucocerebrosidase, as determined according to "Method 1" set forth in Grabowski *et al.*, 1986. 261:8263.

The above-described enzyme assay methods may be advantageously employed for measuring glucocerebrosidase enzymatic activity in various relevant contexts and applications of the present invention.

It will be appreciated that a normal glucocerebrosidase molecule of the present

invention, by virtue of being capable of displaying normal enzymatic activity, will have normally folded 3D structure, in particular with respect to a portion thereof comprising an amino acid residue which, when mutated, alters glucocerebrosidase structure so as to lead to the glucocerebrosidase enzymatic activity deficiency associated with Gaucher disease.

Hence, it will be further appreciated that a normal glucocerebrosidase molecule of the present invention can be advantageously employed for generating a set of structure coordinates defining a 3D structure of a correctly folded glucocerebrosidase molecule, in particular of a correctly folded mutable portion thereof. By virtue of defining such correctly folded 3D structures, such a set of structure coordinates enables generation of a set of structure coordinates defining with optimal accuracy a predicted 3D structure of a mutant glucocerebrosidase molecule, or portion thereof, in particular of a mutant portion thereof.

As described hereinabove, a glucocerebrosidase molecule of the present invention may be characterized by having any of various amino acid sequences, including amino acid sequences displaying any of various partial glycosylation patterns.

Amino acid sequences of normal glucocerebrosidase molecules of the present invention are described hereinbelow.

The glucocerebrosidase molecule may be crystallized using any of various methods, according to the teachings of the present invention.

A glucocerebrosidase molecule of the present invention may be advantageously crystallized according to the method of crystallizing a glucocerebrosidase molecule of the present invention described hereinbelow.

As is described and illustrated in the Examples section below, the present invention provides a crystallized glucocerebrosidase molecule capable of displaying normal enzymatic activity characterized by: (i) an X-ray diffraction capacity enabling generation of a set of structure coordinates defining at 2.0 angstroms an essentially complete 3D structure thereof, including mutable portions thereof; (ii) unit cell dimensions of a = 107.7 angstroms, b = 285.2 angstroms, and c = 91.8 angstroms; (iv) a crystal space group of $C222_1$; and (v) a partially glycosylated amino acid sequence.

As mentioned hereinabove, and as described and illustrated in the Examples section below, while reducing the present invention to practice, the present inventors

used the crystallized glucocerebrosidase molecule of the present invention to generate for the first time a set of structure coordinates defining an essentially complete 3D structure thereof, including that of mutable portions thereof. This set of structure coordinates was used to generate for the first time a computational model representing at atomic resolution such structure. Crystal-derived data of normal and mutant glucocerebrosidase molecules can be used for the first time for computationally identifying a compound or compounds capable of correcting an impaired enzymatic activity of the mutant glucocerebrosidase molecule.

Thus, according to the present invention there is provided a method of identifying a compound capable of correcting an impaired enzymatic activity of a mutant glucocerebrosidase molecule. In a first step, the method is effected by obtaining a set of structure coordinates defining a 3D structure of a crystallized glucocerebrosidase molecule of the present invention capable of normal enzymatic activity, or a portion thereof. In a second step the set of structure coordinates defining the 3D structure of the normal glucocerebrosidase molecule is used to computationally generate a set of structure coordinates defining a predicted 3D structure of the mutant glucocerebrosidase molecule, or a portion thereof. In a third step, the set of structure coordinates defining the predicted 3D structure of the mutant glucocerebrosidase molecule, or portion thereof, is used to computationally identify a compound capable of interacting with the mutant glucocerebrosidase molecule in such a way as to correct the impaired enzymatic activity thereof.

It will be appreciated by one of ordinary skill in the art that a compound capable of interacting with a mutant glucocerebrosidase molecule in such a way as to correct the impaired enzymatic activity thereof, represents a potential Gaucher disease drug, and hence by virtue of enabling identification of such a compound, the method can thereby be used for identifying a candidate Gaucher disease drug.

Preferably, the step of computationally identifying the compound is effected further using the set of structure coordinates defining the 3D structure of the normal glucocerebrosidase molecule of the present invention or the portion thereof.

Following computational identification thereof, the capacity of the compound to correct the impaired enzymatic activity of the mutant glucocerebrosidase molecule is preferably biochemically qualified.

Obtaining the set of structure coordinates defining the 3D structure of the

normal glucocerebrosidase molecule or portion thereof may be effected in various ways.

Preferably, the set of structure coordinates is obtained by subjecting the crystallized glucocerebrosidase molecule of the present invention to crystallographic analysis, as described hereinabove.

Computationally generating the set of structure coordinates defining the predicted 3D structure of the mutant glucocerebrosidase capable of normal molecule, or portion thereof, may be performed in various ways. Preferably, such a set of coordinates is computationally generated according to the method set forth in the Examples section below. Ample guidance for computationally generating such a set of coordinates is provided in the literature of the art, as extensively described hereinbelow.

According to the teachings of the present invention, computationally generating the set of structure coordinates defining the predicted 3D structure of the mutant glucocerebrosidase molecule or portion thereof, is achieved by using as a computationally modified template structure a set of structure coordinates defining a 3D structure of a mutable portion of a normal glucocerebrosidase molecule corresponding to a mutant portion of the mutant glucocerebrosidase molecule. It will be appreciated that since the set of structure coordinates defining a 3D structure of a mutable portion of a normal glucocerebrosidase molecule of the present invention defines an experimentally determined, and hence accurate, 3D atomic structure of the mutable portion of the normal glucocerebrosidase molecule, such a set of coordinates can be utilized for generating a set of structure coordinates defining an optimally accurate 3D atomic structure of a mutant glucocerebrosidase molecule of the present invention.

As described hereinbelow, according to the teachings of the present invention a mutable portion of a normal glucocerebrosidase molecule and a mutant portion of a mutant glucocerebrosidase molecule which correspond to each other preferably comprise sets of amino acid residues which substantially differ from each other only with respect to amino acid residues associated with the impaired glucocerebrosidase enzymatic activity of the mutant glucocerebrosidase molecule. Preferably, such difference will correspond to a single amino acid substitution in the amino acid sequence of the mutant portion of the mutant glucocerebrosidase molecule relative to

the amino acid sequence of the mutable portion of the normal glucocerebrosidase molecule, as described hereinbelow.

The step of using the set of structure coordinates for computationally identifying the compound of the present invention may be effected in various ways. The use of sets of structure coordinates for computationally identifying a compound capable of exerting a desired effect on the functionality of a target enzyme, such as a mutant glucocerebrosidase molecule of the present invention is widely practiced in the art and ample guidance for practicing such computational identification is provided in the literature of the art, as extensively described hereinbelow.

Preferably, computationally identifying the compound of the present invention is effected by computationally qualifying the capacity of a candidate compound to interact with the mutant portion of the mutant glucocerebrosidase molecule, in such a way as to: (i) alter the 3D structure of the mutant portion of the mutant glucocerebrosidase molecule to assume a 3D structure maximally similar to that of the mutable portion of the normal glucocerebrosidase molecule; or (ii) form a complex with the mutant portion of the mutant glucocerebrosidase molecule having a resultant 3D structure maximally similar to that of the mutable portion of the normal glucocerebrosidase molecule. Preferably, in order to optimally computationally qualify the 3D structure of the mutant portion of the glucocerebrosidase molecule, or of a complex thereof with the compound, following interaction of the mutant glucocerebrosidase molecule with the compound, such computational qualifying preferably further comprises using the set of coordinates defining the 3D structure of the mutable portion of the normal glucocerebrosidase molecule as a reference structure to which a 3D structure of the mutant portion of the glucocerebrosidase molecule, or of a complex thereof with the compound, resulting from interaction of the mutant glucocerebrosidase molecule with the compound is compared.

Preferably, the above-described computational processes are reiterated so as to screen a computational library of candidate compound 3D structures so as to identify a compound which interacts with the mutant glucocerebrosidase molecule or mutant portion thereof in such a way as to optimally correct the 3D structure of the mutant glucocerebrosidase molecule or mutant portion thereof, or in such a way as to generate a complex of the compound with the mutant glucocerebrosidase molecule or mutant portion thereof having a desired glucocerebrosidase enzymatic activity when

combined. It will be appreciated that a mutant glucocerebrosidase molecule having a mutant portion whose 3D structure, or of a complex thereof with the compound, is substantially similar to that of the mutable portion of the normal glucocerebrosidase molecule as a result of interacting with the compound will be capable of displaying an increased enzymatic activity relative to such activity in the absence of the compound.

As described hereinabove, since the present invention provides a set of structure coordinates defining at atomic resolution an essentially complete 3D structure of a normal glucocerebrosidase molecule, in particular that of a mutable portion thereof, the method can be used for generating a set of structure coordinates defining with optimal resolution and accuracy a predicted 3D structure of essentially any mutable portion of the normal glucocerebrosidase molecule. It will be appreciated that a set of structure coordinates defining a predicted 3D structure of essentially any mutable portion of a normal glucocerebrosidase molecule can be used for optimally qualifying a capacity of a candidate compound to interact with a mutant portion of a mutant glucocerebrosidase molecule in such a way as to correct a 3D structure of a mutant portion thereof, and hence for optimally correcting the impaired enzymatic activity of the mutant glucocerebrosidase molecule. Hence, as described hereinbelow, the method of identifying the compound of the present invention can be used for identifying an optimal Gaucher disease drug.

Depending on the desired mutant glucocerebrosidase molecule specificity of the compound of the present invention, the mutant glucocerebrosidase molecule may have any of various amino acid sequences. Preferably, the mutant glucocerebrosidase molecule comprises a mutant portion corresponding to the desired mutant glucocerebrosidase molecule specificity of the compound of the present invention.

Amino acid sequences of mutant glucocerebrosidase molecules of the present invention, and specific applications thereof are described hereinbelow.

It will be appreciated that by virtue of enabling computational screening of libraries of compounds having essentially any of various chemical, biological and/or physical characteristics, the method enables identification of a compound capable of displaying optimal *in-vivo* pharmacokinetics, optimally low immunogenicity, and optimal effectiveness relative to all prior art compounds capable of correcting impaired glucocerebrosidase enzymatic activity of a mutant glucocerebrosidase molecule, such as DNJ-based compounds, or prior art compounds capable of

providing exogenous glucocerebrosidase activity, such as Cerezyme®. Hence, the method can be utilized to identify an optimal Gaucher disease drug.

For optimal effectiveness as a Gaucher disease drug *in-vivo*, the compound of the present invention should be as small as possible while still being capable of correcting the impaired enzymatic activity of the mutant glucocerebrosidase molecule since, keeping all other parameters equal, the smaller a molecule, the better its overall biodistribution capacity will generally be. An optimally small compound of the present invention will circumvent the limitations of prior art Gaucher disease drugs such as Cerezyme® whose active ingredient is a full-sized protein.

The molecule is preferably a non-polypeptidic compound. A non-polypeptidic compound will avoid the immunogenicity drawbacks of prior art Gaucher disease drugs such as Cerezyme® whose active ingredient is a protein.

It will be appreciated that, by virtue of its capacity for correcting impaired endogenous glucocerebrosidase activity, the compound of the present invention is distinctly advantageous over prior art Gaucher disease drugs since, unlike prior art compounds such as Cerezyme®, it affords therapeutic glucocerebrosidase activity with optimal spatial (cell type/subcellular location), temporal, and activity level regulation. Evidence for the feasibility of using a compound for correcting an impaired enzymatic activity of a mutant glucocerebrosidase molecule associated with Gaucher disease has been provided in the prior art (Leslie AGW., 1992. Newsletter on Protein Crystallography 26). Examples of suitable chemical structure databases for performing the above-described screening are described hereinbelow.

The compound of the present invention can be used *per se* or it can be formulated as the active ingredient of a pharmaceutical composition comprising suitable carriers and/or diluents, and an effective concentration of the compound of the present invention so as to be suitable for therapeutically correcting the impaired enzymatic activity of the mutant glucocerebrosidase molecule when administered to a subject in need thereof, in particular a subject having Gaucher disease.

One of ordinary skill in the art, such as for example, a pharmacologist, a veterinarian, or a physician would possess the expertise required for correctly formulating such a pharmaceutical composition for effective disease treatment.

Biochemical qualification of the capacity of the compound to correct the impaired enzymatic activity of the mutant glucocerebrosidase molecule:

As described hereinabove, following computational identification thereof, the capacity of the compound to correct the impaired enzymatic activity of the mutant glucocerebrosidase molecule is preferably biochemically qualified.

Preferably, biochemically qualifying the capacity of the compound to correct the impaired enzymatic activity of the mutant glucocerebrosidase molecule is effected by measuring the specific glucocerebrosidase enzymatic activity of the mutant glucocerebrosidase molecule in the absence or presence of the compound using a standard glucocerebrosidase activity assay. A compound identified capable of increasing, or optimally increasing, the specific glucocerebrosidase enzymatic activity of the mutant glucocerebrosidase molecule constitutes a candidate Gaucher disease drug, or a candidate optimal Gaucher disease drug, respectively. The specific enzymatic activity of the mutant glucocerebrosidase molecule may be advantageously measured in the presence of varying concentrations of the compound so as to enable identification of concentrations of the compound needed for increasing the specific enzymatic activity of the mutant glucocerebrosidase molecule to an optimal level.

Any of various glucocerebrosidase enzymatic activity assays may be employed.

For example, the specific glucocerebrosidase enzymatic activity of the mutant glucocerebrosidase molecule may be determined using an enzyme assay using as an enzyme substrate either glucosylceramidase, the natural substrate of glucocerebrosidase, or a synthetic glucocerebrosidase substrate, preferably a synthetic beta-glycoside substrate of glucocerebrosidase. Ample art guidance for practicing such methods is provided hereinabove.

Preferably, the compound of the present invention is selected capable of sufficiently correcting the impaired enzymatic activity of the mutant glucocerebrosidase molecule so as to have utility for treating Gaucher disease. Ample guidance regarding normal glucocerebrosidase enzymatic activity levels is provided hereinabove.

As described hereinabove, while reducing the present invention to practice the present inventors produced a computing platform capable of generating a model representing an essentially complete 3D atomic structure of a glucocerebrosidase

molecule.

Thus, according to still another aspect of the present invention there is provided a computing platform capable of generating a model representing a 3D structure of a glucocerebrosidase molecule or a portion thereof. The computing platform comprises a data-storage device storing data comprising a set of structure coordinates defining the 3D structure of the glucocerebrosidase molecule or the portion thereof. The computing platform further comprises a processing unit for generating the model representing the 3D structure from the data stored in the data-storage device.

Examples of suitable platforms and software applications for running such platforms are provided hereinbelow.

The computing platform of the present invention can be exploited in various ways.

As described hereinabove, the computing platform of the present invention can be used as described hereinabove for identifying a compound capable of optimally correcting an impaired enzymatic activity of essentially any mutant glucocerebrosidase molecule, and hence for identifying a candidate Gaucher disease drug. It will be appreciated that the computing platform can further be used for optimally elucidating the structural and functional characteristics of normal and mutant glucocerebrosidase molecules, and thereby for generating novel information having significant utility in fields such as medical diagnostics, therapeutics, pharmacology, and applied research. This is abundantly demonstrated by the wealth of novel structural/functional features of normal and mutant glucocerebrosidase molecules which can now be described for the first time using the computing platform of the present invention.

Depending on the application and purpose, the computing platform may be capable of generating a model representing a 3D structure of essentially any normal glucocerebrosidase molecule or portion thereof and/or may be capable of generating a model representing a 3D structure of essentially any mutant glucocerebrosidase molecule or portion thereof.

Normal glucocerebrosidase molecules and mutant glucocerebrosidase molecules of the present invention are described hereinbelow.

Preferably, a computing platform capable of generating a model representing a

3D structure of a portion of a normal glucocerebrosidase molecule of the present invention is capable of generating a model representing a 3D structure of a mutable portion of a normal glucocerebrosidase molecule of the present invention.

Preferably, a computing platform capable of generating a model representing a 3D structure of a portion of a mutant glucocerebrosidase molecule of the present invention is capable of generating a model representing a 3D structure of a mutant portion of a mutable glucocerebrosidase molecule of the present invention.

As described hereinabove, a computing platform capable of generating a model representing a 3D structure of a mutable portion of a normal glucocerebrosidase molecule of the present invention can be used for computationally generating a set of structure coordinates defining a predicted 3D structure of a mutant glucocerebrosidase of the present invention, and for facilitating computational identification of the compound of the present invention.

As described hereinabove, a computing platform capable of generating a model representing a 3D structure of a mutant portion of a mutant glucocerebrosidase molecule of the present invention can be used for identifying the compound of the present invention.

Amino acid sequence of normal glucocerebrosidase molecule of present invention:

As described hereinabove, a normal glucocerebrosidase molecule of the present invention may have any of various amino acid sequences, and thereby may comprise any of various mutable portions, depending on the application and purpose.

The glucocerebrosidase molecule of the present invention may be characterized by a wild-type amino acid sequence (SEQ ID NO: 8), or by an amino acid sequence in which Arg495 of the wild-type amino acid sequence is substituted with a His residue (SEQ ID NO: 1).

Preferably, the amino acid sequence of the normal glucocerebrosidase molecule of the present invention is composed of 497 amino acid residues.

Preferably, the amino acid sequence of a normal glucocerebrosidase molecule of the present invention is set forth in SEQ ID NO: 1.

A glucocerebrosidase molecule having the amino acid sequence set forth in SEQ ID NO: 1 is widely understood in the art as being capable of displaying wild-type enzymatic activity. For example, as described above, the standard treatment for

Gaucher disease is enzyme replacement therapy with the commercial drug Cerezyme® (Genzyme Corporation, Cambridge, MA, USA) whose active ingredient is a variant of human glucocerebrosidase having the amino acid sequence set forth in SEQ ID NO: 1, and exhibiting an enzymatic activity identical to that of wild-type human glucocerebrosidase (for example, refer to Grabowski GA. *et al.*, 1995. Ann Intern Med. 122:33-9). Substitution of Arg495 to His in Cerezyme® has been documented not to affect the catalytic functions, safety or therapeutic effectiveness of Cerezyme (Genzyme Corporation, on file; Grabowski et al., 1995. Ann Intern Med. 122: 33-39; Grace et al., 1993. J Biol Chem. 265:2283-2291).

Hence, it will be appreciated that a 3D structure of Cerezyme® is essentially identical to that of wild-type human glucocerebrosidase, in particular with respect to mutable portions thereof.

Portion of normal glucocerebrosidase molecule of present invention:

As described hereinabove, a normal glucocerebrosidase molecule of the present invention may be characterized by any of various portions depending on the application and purpose.

Preferably, a portion of a normal glucocerebrosidase molecule comprises a set of amino acid residues of the amino acid sequence of the normal glucocerebrosidase molecule having at least one atom positioned within 10 angstroms of at least one atom of a reference amino acid residue of the normal glucocerebrosidase molecule, as described in Example 2 of the Examples section below.

Preferably, a mutable portion of a normal glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues comprises an amino acid residue located at position 370, 394, 409, 444, 463, and/or 496 of the amino acid sequence of the normal glucocerebrosidase molecule.

As described in the Examples section below, mutation of an amino acid residue located at position 370, 394, 409, 444, 463, or 496 of the amino acid sequence of a glucocerebrosidase molecule, such as a glucocerebrosidase molecule having the amino acid sequence set forth in SEQ ID NO: 1 or 8, in particular a mutation characterized by a Ser residue at position 370, a Leu residue at position 394, a His residue at position 409, a Pro residue at position 444, or a His residue at position 496, is associated with one of the most common forms of Gaucher disease.

Preferably, the mutable portion comprising the amino acid residue at position

370 of the amino acid sequence of the normal glucocerebrosidase molecule comprises the amino acid residues at positions 76, 81, 285, 312, 314, 320, 324, 325, 336, 364–378, 423, and 433 of the amino acid sequence of the normal glucocerebrosidase molecule.

Preferably, the mutable portion comprising the amino acid residue at position 394 of the amino acid sequence of the normal glucocerebrosidase molecule comprises the amino acid residues at positions 244–247, and 390–397 of the amino acid sequence of the normal glucocerebrosidase molecule.

Preferably, the mutable portion comprising the amino acid residue at position 409 of the amino acid sequence of the normal glucocerebrosidase molecule comprises the amino acid residues at positions 20, 21, 95–100, and 404–411 of the amino acid sequence of the normal glucocerebrosidase molecule.

Preferably, the mutable portion comprising the amino acid residue at position 444 of the amino acid sequence of the normal glucocerebrosidase molecule comprises the amino acid residues at positions 65–67, 440–447, 460–464, 468, and 469 of the amino acid sequence of the normal glucocerebrosidase molecule.

Preferably, the mutable portion comprising the amino acid residue at position 463 of the amino acid sequence of the normal glucocerebrosidase molecule comprises the amino acid residues at positions 360–366, 443–446, 460–467, and 484–89 of the amino acid sequence of the normal glucocerebrosidase molecule.

Preferably, the mutable portion comprising the amino acid residue at position 496 of the amino acid sequence of the normal glucocerebrosidase molecule comprises the amino acid residues at positions 33–35, 69, 71, 450–456, 474–478, and 493–497 of the amino acid sequence of the normal glucocerebrosidase molecule.

3D structure of normal glucocerebrosidase molecule of present invention:

As is described hereinabove, a normal glucocerebrosidase molecule of the present invention, or a portion thereof, may be characterized by any of various 3D structures, depending on the application and purpose.

Preferably, the normal glucocerebrosidase molecule of the present invention is characterized by a 3D structure defined by the set of structure coordinates set forth in Table 4.

Refer to enclosed CD-ROM for Tables 4-22.

As is described in the Examples section which follows, the two sets of

structure coordinates corresponding to atom coordinates 1-3929 and 3930-7859 set forth in Table 4 define at atomic resolution two essentially complete, and essentially identical, 3D structures of two distinct co-crystallized glucocerebrosidase molecules of the present invention capable of normal enzymatic activity. As is described in Example 1 of the Examples section which follows, these two structures represent the structures of the two monomers forming the asymmetric unit of crystallized glucocerebrosidase molecules of the present invention capable of normal enzymatic activity generated while reducing the present invention to practice.

While a normal glucocerebrosidase molecule of the present invention whose 3D structure is defined by a set of structure coordinates set forth in Table 4 may have a 3D structure defined by either the set of structure coordinates corresponding to atom coordinates 1-3929 or the set of structure coordinates corresponding to atom coordinates 3930-7859 set forth in Table 4, its 3D structure is preferably defined by the set of structure coordinates corresponding to atom coordinates 1-3929 set forth in Table 4.

As is described and illustrated in Example 1 of the Examples section which follows, the set of structure coordinates corresponding to atom coordinates 1-3929 set forth in Table 4 defines an essentially complete 3D atomic structure of a normal glucocerebrosidase molecule of the present invention.

Hence, as described hereinabove, the set of structure coordinates set forth in Table 4 can be utilized for generating a model representing a 3D structure of a normal glucocerebrosidase molecule of the present invention, or a portion thereof.

3D structure of portion of normal glucocerebrosidase molecule of the present invention:

As described hereinabove, a portion of a normal glucocerebrosidase molecule of the present invention may be characterized by any of various 3D structures.

Preferably, a mutable portion of a normal glucocerebrosidase molecule of the present invention whose structure is set forth in Table 4 has a 3D structure defined by the set of structure coordinates set forth in Table 5, 6, 7, 8, 9, and/or 10.

As is described and illustrated in Example 2 of the Examples section which follows, the set of structure coordinates set forth in Table 5 defines a 3D structure of a mutable portion of a normal glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues where such

portion comprises amino acid residues at positions 76, 81, 285, 312, 314, 320, 324, 325, 336, 364–378, 423, and 433 of the amino acid sequence of the normal glucocerebrosidase molecule. Hence, as described hereinbelow, this set of structure coordinates can be utilized for generating a computational model representing at atomic resolution an essentially complete 3D structure of such a portion of such a glucocerebrosidase molecule.

As is described and illustrated in Example 2 of the Examples section which follows, the set of structure coordinates set forth in Table 6 defines at optimally high resolution an essentially complete 3D structure of a portion of a normal glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues where such portion comprises amino acid residues at positions 244–247, and 390–397 of the amino acid sequence of the normal glucocerebrosidase molecule. Hence, as described hereinbelow, this set of structure coordinates can be utilized for generating a computational model representing at atomic resolution an essentially complete 3D structure of such a portion of such a glucocerebrosidase molecule.

As is described and illustrated in Example 2 of the Examples section which follows, the set of structure coordinates set forth in Table 7 defines at optimally high resolution an essentially complete 3D structure of a portion of a normal glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues where such portion comprises amino acid residues at positions 20, 21, 95–100, and 404–411 of the amino acid sequence of the normal glucocerebrosidase molecule. Hence, as described hereinbelow, this set of structure coordinates can be utilized for generating a computational model representing at atomic resolution an essentially complete 3D structure of such a portion of such a glucocerebrosidase molecule.

As is described and illustrated in Example 2 of the Examples section which follows, the set of structure coordinates set forth in Table 8 defines at optimally high resolution an essentially complete 3D structure of a portion of a normal glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues where such portion comprises amino acid residues at positions 65–67, 440–447, 460–464, 468, and 469 of the amino acid sequence of the normal glucocerebrosidase molecule. Hence, as described

hereinbelow, this set of structure coordinates can be utilized for generating a computational model representing at atomic resolution an essentially complete 3D structure of such a portion of such a glucocerebrosidase molecule.

As is described and illustrated in Example 2 of the Examples section which follows, the set of structure coordinates set forth in Table 9 defines at optimally high resolution an essentially complete 3D structure of a portion of a normal glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues where such portion comprises amino acid residues at positions 360–366, 443–446, 460–467, and 484–89 of the amino acid sequence of the normal glucocerebrosidase molecule. Hence, as described hereinbelow, this set of structure coordinates can be utilized for generating a computational model representing at atomic resolution an essentially complete 3D structure of such a portion of such a glucocerebrosidase molecule.

As is described and illustrated in Example 2 of the Examples section which follows, the set of structure coordinates set forth in Table 10 defines at optimally high resolution an essentially complete 3D structure of a portion of a normal glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues where such portion comprises amino acid residues at positions 33–35, 69, 71, 450–456, 474–478, and 493–497 of the amino acid sequence of the normal glucocerebrosidase molecule. Hence, as described hereinbelow, this set of structure coordinates can be utilized for generating a computational model representing at atomic resolution an essentially complete 3D structure of such a portion of such a glucocerebrosidase molecule.

It will be appreciated that mutant glucocerebrosidase molecules having amino acid sequences composed of 497 amino acid residues where such portions have a Ser residue at position 370, a Leu residue at position 394, a His residue at position 409, a Pro residue at position 444, and/or a His residue at position 496 of such amino acid sequences represent the most common mutant glucocerebrosidase molecules associated with Gaucher disease (refer, for example, to Table 3 of the Examples section). It will be further appreciated that the sets of structure coordinates set forth in Tables 5-10 define at atomic resolution essentially complete 3D structures of mutable portions of a normal glucocerebrosidase molecule having an amino acid sequence composed of 497 amino acid residues where such portions comprise the amino acid

residue located at position 370, 394, 409, 444, 463, or 496 of the amino acid sequence of such a glucocerebrosidase molecule, respectively. Hence, the set of structure coordinates set forth in Tables 5-10 can respectively be used for the first time for generating computing platforms capable of generating models representing at atomic resolution essentially complete 3D structures of mutable portions comprising amino acid residues located at position 370, 394, 409, 444, 463, and/or 496 of the amino acid sequence of such a normal glucocerebrosidase molecule. As such, as described hereinbelow and in the Examples section below, the set of structure coordinates set forth in Tables 5-10 can be used for the first time for facilitating computational identification of drugs optimally suitable for treating the most prevalent forms of Gaucher disease.

Amino acid sequence of mutant glucocerebrosidase molecule of present invention:

As described hereinabove, a mutant glucocerebrosidase molecule of the present invention may have any of various amino acid sequences.

Preferably, the amino acid sequence of the mutant glucocerebrosidase molecule is composed of 497 amino acid residues.

Preferably, the amino acid sequence of the glucocerebrosidase molecule is set forth in SEQ ID NO: 2, 3, 4, 5, 6, or 7, depending on the mutant glucocerebrosidase molecule of the present invention of interest.

It will be appreciated that the amino acid sequences set forth in SEQ ID NOs: 2-7 are the amino acid sequences of mutant glucocerebrosidase molecules having an amino acid sequence composed of 497 amino acid residues having an amino acid sequence comprising the mutant amino acid residue Ser at position 370, Leu at position 394, His at position 409, Pro at position 444, and His at position 496, respectively. As described hereinabove, such residues at such positions correspond to the most common single amino acid substitution mutations associated with Gaucher disease.

Alternately, the mutant glucocerebrosidase of the present invention may have the amino acid sequence set forth in SEQ ID NO: 9, 10, 11, 12, 13, or 14.

It will be appreciated that the amino acid sequences set forth in SEQ ID NOs: 9-14 are the amino acid sequences of naturally occurring mutant glucocerebrosidase molecules comprising the mutated amino acid residue Ser at position 370, Leu at

position 394, His at position 409, Pro at position 444, and His at position 496, respectively, and that, as described hereinabove, such mutant glucocerebrosidase molecules comprise the single amino acid substitution mutations most commonly associated with Gaucher disease.

Other naturally occurring mutations of human glucocerebrosidase which may be associated with less frequent variants of Gaucher disease are described in Figure 1d of the Examples section which follows, and in the literature of the art [for example, refer to Beutler E. and Grabowski GA., in: "The Metabolic and Molecular Bases of Inherited Disease", Scriver CR. *et al.* (eds.), McGraw-Hill Inc., pp. 3635-3668 (2001)].

Portion of mutant glucocerebrosidase molecule of present invention:

As described hereinabove, a mutant glucocerebrosidase molecule of the present invention may be characterized by any of various portions depending on the application and purpose.

Preferably, a portion of a mutant glucocerebrosidase molecule comprises a set of amino acid residues of the amino acid sequence of the mutant glucocerebrosidase molecule having at least one atom positioned within 10 angstroms of at least one atom of a reference amino acid residue of the amino acid sequence of the mutant glucocerebrosidase molecule, as described in Example 2 of the Examples section below.

Preferably, a mutant portion of a mutant glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues comprises a mutated amino acid residue located at position 370, 394, 409, 444, 463, and/or 496 of the amino acid sequence of the mutant glucocerebrosidase molecule.

Preferably, the mutated amino acid residue is a Ser residue at position 370, a Leu residue at position 394, a His residue at position 409, a Pro residue at position 444, and/or a His residue at position 496.

As described hereinabove, such mutations at such positions in glucocerebrosidase molecules having amino acid sequences composed of 497 amino acid residues represent the single amino acid substitution mutations associated with the most common forms of Gaucher disease.

Preferably, the mutant portion comprising the amino acid residue at position

370 of the amino acid sequence of the mutant glucocerebrosidase molecule comprises the amino acid residues at positions 76, 81, 285, 312, 314, 320, 324, 325, 336, 364–378, 423, and 433 of the amino acid sequence of the mutant glucocerebrosidase molecule.

Preferably, the mutant portion comprising the amino acid residue at position 394 of the amino acid sequence of the mutant glucocerebrosidase molecule comprises the amino acid residues at positions 244–247, and 390–397 of the amino acid sequence of the mutant glucocerebrosidase molecule.

Preferably, the mutant portion comprising the amino acid residue at position 409 of the amino acid sequence of the mutant glucocerebrosidase molecule comprises the amino acid residues at positions 20, 21, 95–100, and 404–411 of the amino acid sequence of the mutant glucocerebrosidase molecule.

Preferably, the mutant portion comprising the amino acid residue at position 444 of the amino acid sequence of the mutant glucocerebrosidase molecule comprises the amino acid residues at positions 65–67, 440–447, 460–464, 468, and 469 of the amino acid sequence of the mutant glucocerebrosidase molecule.

Preferably, the mutant portion comprising the amino acid residue at position 463 of the amino acid sequence of the mutant glucocerebrosidase molecule comprises the amino acid residues at positions 360–366, 443–446, 460–467, and 484–89 of the amino acid sequence of the mutant glucocerebrosidase molecule.

Preferably, the mutant portion comprising the amino acid residue at position 496 of the amino acid sequence of the mutant glucocerebrosidase molecule comprises the amino acid residues at positions 33–35, 69, 71, 450–456, 474–478, and 493–497 of the amino acid sequence of the mutant glucocerebrosidase molecule.

3D structure of mutant glucocerebrosidase molecule of present invention:

As is described hereinabove, a mutant glucocerebrosidase molecule of the present invention, or a portion thereof, may be characterized by any of various 3D structures, depending on the application and purpose.

Preferably, the set of structure coordinates defining the structure of a mutant glucocerebrosidase molecule of the present invention defines the 3D structure of a crystallized form of such a mutant glucocerebrosidase of the present invention.

Preferably, the mutant glucocerebrosidase molecule of the present invention is characterized by a 3D structure defined by the set of structure coordinates set forth in

Table 11, 13, 15, 17, 19, or 21.

As is described in the Examples section which follows, the set of structure coordinates set forth in Table 11, 13, 15, 17, 19, or 21 defines at atomic resolution an essentially complete predicted 3D structure of a mutant glucocerebrosidase molecule having the amino acid sequence set forth in SEQ ID NO: 2, 3, 4, 5, 6, or 7, respectively.

Hence, as described hereinabove, the set of structure coordinates set forth in Table 11, 13, 15, 17, 19, or 21 can be utilized for generating a model representing at atomic resolution an essentially complete predicted 3D structure of a mutant glucocerebrosidase molecule of the present invention having the amino acid sequence set forth in SEQ ID NO: 2, 3, 4, 5, 6, or 7, respectively.

3D structure of portion of mutant glucocerebrosidase molecule of the present invention:

As described hereinabove, a portion of a mutant glucocerebrosidase molecule of the present invention may be characterized by any of various 3D structures, depending on the mutant glucocerebrosidase molecule of interest.

Preferably, a mutant portion of a mutant glucocerebrosidase molecule of the present invention whose structure is set forth in Table 11, 13, 15, 17, 19, or 21 has a 3D structure defined by the set of structure coordinates set forth in Table 12, 14, 16, 18, 20 or 22, respectively.

As is described and illustrated in Example 2 of the Examples section which follows, the set of structure coordinates set forth in Table 12 defines a 3D structure of a mutant portion of a mutant glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues where such portion comprises amino acid residues at positions 76, 81, 285, 312, 314, 320, 324, 325, 336, 364–378, 423, and 433 of the amino acid sequence of the mutant glucocerebrosidase molecule. Hence, as described hereinbelow, this set of structure coordinates can be utilized for generating a computational model representing at atomic resolution an essentially complete predicted 3D structure of such a portion of such a glucocerebrosidase molecule.

As is described and illustrated in Example 2 of the Examples section which follows, the set of structure coordinates set forth in Table 14 defines at optimally high resolution an essentially complete 3D structure of a mutant portion of a mutant

glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues where such portion comprises amino acid residues at positions 244–247, and 390–397 of the amino acid sequence of the mutant glucocerebrosidase molecule. Hence, as described hereinbelow, this set of structure coordinates can be utilized for generating a computational model representing at atomic resolution an essentially complete predicted 3D structure of such a portion of such a glucocerebrosidase molecule.

As is described and illustrated in Example 2 of the Examples section which follows, the set of structure coordinates set forth in Table 16 defines at optimally high resolution an essentially complete 3D structure of a portion of a mutant glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues where such portion comprises amino acid residues at positions 20, 21, 95–100, and 404–411 of the amino acid sequence of the mutant glucocerebrosidase molecule. Hence, as described hereinbelow, this set of structure coordinates can be utilized for generating a computational model representing at atomic resolution the essentially complete predicted 3D structure of such a portion of such a glucocerebrosidase molecule.

As is described and illustrated in Example 2 of the Examples section which follows, the set of structure coordinates set forth in Table 18 defines at optimally high resolution an essentially complete 3D structure of a portion of a mutant glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues where such portion comprises amino acid residues at positions 65–67, 440–447, 460–464, 468, and 469 of the amino acid sequence of the mutant glucocerebrosidase molecule. Hence, as described hereinbelow, this set of structure coordinates can be utilized for generating a computational model representing at atomic resolution the essentially complete predicted 3D structure of such a portion of such a glucocerebrosidase molecule.

As is described and illustrated in Example 2 of the Examples section which follows, the set of structure coordinates set forth in Table 20 defines at optimally high resolution an essentially complete 3D structure of a portion of a mutant glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues where such portion comprises amino acid residues at positions 360–366, 443–446, 460–467, and 484–89 of the amino acid

sequence of the mutant glucocerebrosidase molecule. Hence, as described hereinbelow, this set of structure coordinates can be utilized for generating a computational model representing at atomic resolution the essentially complete predicted 3D structure of such a portion of such a glucocerebrosidase molecule.

As is described and illustrated in Example 2 of the Examples section which follows, the set of structure coordinates set forth in Table 22 defines at optimally high resolution an essentially complete 3D structure of a portion of a mutant glucocerebrosidase molecule of the present invention having an amino acid sequence composed of 497 amino acid residues where such portion comprises amino acid residues at positions 33–35, 69, 71, 450–456, 474–478, and 493–497 of the amino acid sequence of the mutant glucocerebrosidase molecule. Hence, as described hereinbelow, this set of structure coordinates can be utilized for generating a computational model representing at atomic resolution the essentially complete predicted 3D structure of such a portion of such a glucocerebrosidase molecule.

As described hereinabove, mutant glucocerebrosidase molecules having an amino acid sequence composed of 497 amino acid residues comprising a Ser residue at position 370, a Leu residue at position 394, a His residue at position 409, a Pro residue at position 444, and/or a His residue at position 496 represent the most common mutant glucocerebrosidase molecules associated with Gaucher disease. It will be appreciated that the set of structure coordinates set forth in Table 12, 14, 16, 18, 20 or 22 respectively defines at atomic resolution an essentially complete 3D structure of a mutant portion of such a glucocerebrosidase molecule where such portion comprises the amino acid residue located at position 370, 394, 409, 444, 463, and/or 496 of the amino acid sequence of such a glucocerebrosidase molecule. Hence, the set of structure coordinates set forth in Table 12, 14, 16, 18, 20 or 22 can respectively be used, for the first time, for generating a computer-generated model representing at atomic resolution an essentially complete predicted 3D structure of such a mutable portion of such a mutant glucocerebrosidase molecule. As such, as described hereinbelow and in the Examples section below, the set of structure coordinates set forth in Table 12, 14, 16, 18, 20 and/or 22 can be used for the first time for facilitating computational identification of drugs optimally suitable for treating the most prevalent forms of Gaucher disease.

Glycosylation:

As described hereinabove, the glucocerebrosidase molecule of the present invention may have an amino acid sequence comprising any of various glycosylation patterns.

Preferably, the glucocerebrosidase molecule of the present invention has a partially glycosylated amino acid sequence.

As used herein, the phrase "partially glycosylated amino acid sequence" refers to an amino acid sequence of a glucocerebrosidase molecule of the present invention having a glycosylation pattern in which at least one N-linked glycosylation consensus sequence thereof which is naturally glycosylated is non-glycosylated. It will be understood by the ordinarily skilled artisan that N-linked glycosylation consensus sequences of the amino acid sequence of glucocerebrosidase which are normally glycosylated are located at positions 19, 59, 146, and 270 of the amino acid sequence (for example, refer to Berg-Fussman A. *et al.*, 1993. J Biol Chem. 268:14861-14866).

Preferably, the partially glycosylated amino acid sequence comprises a sugar moiety attached to the first N-linked glycosylation consensus sequence thereof.

As used herein, the term "first" when referring to an N-linked glycosylation consensus sequence of an amino acid sequence corresponds to the N-linked glycosylation consensus sequence of the amino acid sequence located closest to the amino-terminal of the amino acid sequence. For example, in the amino acid sequence set forth in SEQ ID NO: 1 or 8, the first N-linked glycosylation consensus sequence is located at position 19 of the amino acid sequence.

As is widely understood in the art, an N-linked glycosylation consensus sequence corresponds to the amino acid sequence Asn-Xaa-(Ser/Thr) (SEQ ID NO: 15), where the invariant Asn may be glycosylated (Berg-Fussman A. *et al.*, 1993. J Biol Chem. 268:14861-14866). Such an N-linked glycosylation consensus sequence may also be referred to as a "sequon" in the art.

The sugar moiety attached to the first N-linked glycosylation consensus sequence of the amino acid sequence of the crystallized glucocerebrosidase molecule may consist of various numbers and/or types of sugars.

Preferably, the sugar moiety is a monosaccharide or a disaccharide.

Preferably, the sugar is N-acetylglucosamine.

Preferably, the sugar moiety is the only sugar moiety attached to an N-linked

glycosylation consensus sequence of the partially glycosylated amino acid sequence.

An amino acid sequence of a glucocerebrosidase molecule having as its sole glycosylation of an N-linked glycosylation consensus sequence a glycosylation or partial glycosylation of its first N-linked glycosylation consensus sequence is widely understood in the art being capable of displaying normal enzymatic activity (for example, refer to Berg-Fussman A. *et al.*, 1993. J Biol Chem. 268:14861-14866).

As described hereinabove, while reducing the present invention to practice, the present inventors unexpectedly devised a method of crystallizing a glucocerebrosidase molecule having optimal X-ray diffraction capacity.

Thus, according to the present invention there is provided a method of crystallizing a glucocerebrosidase molecule. The method is effected by partially deglycosylating the glucocerebrosidase molecule to thereby generate a partially glycosylated glucocerebrosidase molecule, and subjecting the partially glycosylated glucocerebrosidase molecule to crystallization-inducing conditions.

As described hereinabove, the method can be used to generate a crystallized glucocerebrosidase molecule capable of diffracting X-rays so as to enable for the first time generation of a set of structure coordinates defining at atomic resolution an essentially complete 3D structure of a normal glucocerebrosidase molecule of the present invention.

Partially deglycosylating the glucocerebrosidase molecule may be effected in various ways, for example using any of various glycosidases.

Preferably, partially deglycosylating the glucocerebrosidase molecule is effected by treating the glucocerebrosidase molecule with N-glycosidase F, preferably according to the protocol set forth in Example 1 of the Examples section below.

Preferably, the step of partially deglycosylating the glucocerebrosidase molecule is effected so as to leave a monosaccharide or a disaccharide of the present invention attached to the first N-linked glycosylation consensus sequence of the amino acid sequence of the glucocerebrosidase molecule.

Preferably, the step of partially deglycosylating the glucocerebrosidase molecule is further effected so as to effectively fully deglycosylate all glycosylated N-linked glycosylation consensus sequences of the amino acid sequence other than the first N-linked glycosylation consensus sequence.

Without being bound to a paradigm, the present inventors are of the opinion

that the above-described deglycosylation method enabled for the first time crystallization of a glucocerebrosidase molecule having the X-ray diffraction capacity of the crystallized glucocerebrosidase molecule of the present invention by virtue of allowing optimal packing of glucocerebrosidase molecules in the crystal.

Preferably, the crystallization-inducing conditions comprise inducing evaporation of a crystallization solution containing the partially glycosylated glucocerebrosidase molecule at a concentration of about 5 mg/ml, a buffer, a sodium salt, an ammonium salt, a sulfate salt, a chaotropic compound, a potassium salt, and/or a chloride ion. Most preferably, the crystallization-inducing conditions comprise inducing evaporation of a crystallization solution containing a maximal number of the aforementioned components as well as the partially glycosylated glucocerebrosidase molecule.

Inducing evaporation of the crystallization solution may be effected in various ways.

Preferably, inducing evaporation of the crystallization solution is effected by inducing evaporation of a hanging drop thereof.

Preferably, inducing such evaporation is effected at a temperature of about 22 degrees centigrade.

Typically when inducing evaporation of a hanging drop of crystallization solution, a small drop of crystallization mixture containing a macromolecule to be crystallized is placed on a cover slip or glass plate which is inverted over a well of equilibration solution such that the cover slip or glass plate forms a seal over the well. The equilibration solution is initially at a lower volatile component vapor pressure than the crystallization mixture so that evaporation of the volatile component from the crystallization mixture to the equilibration mixture progresses at a rate fixed by the difference in the vapor pressures therebetween and by the distance between the crystallization mixture and the equilibration solution. Thus, as evaporation proceeds, the crystallization mixture becomes supersaturated with the macromolecule to be crystallized and, under the appropriate crystallization mixture conditions-including pH, solute composition and/or concentration, and temperature-crystallization occurs.

Preferably, the buffer is a Zwitterionic buffer or an acetate buffer.

Preferably, the Zwitterionic buffer is 2-morpholinoethanesulfonic acid buffer.

Preferably, the concentration of the Zwitterionic buffer in the crystallization

solution is about 0.5 millimolar.

Preferably, the pH of the Zwitterionic buffer is about 6.6.

Preferably, the acetate buffer is sodium acetate buffer.

Preferably, the concentration of the acetate buffer in the crystallization solution is about 0.05 molar.

Preferably, the pH of the acetate buffer is about 4.6.

Preferably, the sodium salt is sodium chloride.

Preferably, the concentration of the sodium salt in the crystallization solution is about 0.05 molar.

Preferably, the ammonium salt is ammonium sulfate.

Preferably, the concentration of the ammonium salt in the crystallization solution is about 0.5 molar.

Preferably, the concentration of the sulfate salt in the crystallization solution is about 0.5 molar.

Preferably, the chaotropic compound is guanidine hydrochloride.

Preferably, the concentration of the chaotropic compound in the crystallization solution is about 0.085 molar.

Preferably, the potassium salt is potassium chloride.

Preferably, the concentration of the potassium salt in the crystallization solution is about 0.01 molar.

Preferably, the concentration of the chloride ion in the crystallization solution is about 0.06 molar.

The crystallization solution may have any of various pH's.

Preferably, the crystallization solution has a pH of about 4.6.

As described hereinabove, while reducing the present invention to practice the present inventors generated a set of structure coordinates defining a 3D structure of a glucocerebrosidase molecule of the present invention or a portion thereof. It will be appreciated that such a set of structure coordinates can be used to produce a computer-readable medium comprising, in a retrievable format, data including such a set of structure coordinates.

Thus, according to the present invention there is provided a computer-readable medium comprising, in a retrievable format, data including a set of structure coordinates defining a 3D structure of a glucocerebrosidase molecule or a portion

thereof, wherein the set of structure coordinates defines the 3D structure at a resolution of 2.9 angstroms or higher, and/or wherein the amino acid sequence of the glucocerebrosidase molecule is partially glycosylated.

Such a computer-readable medium can be used for producing a computing platform capable of generating a model representing a 3D structure of a glucocerebrosidase molecule of the present invention or a portion thereof.

Thus, according to the present invention there is provided a computer generated model representing a 3D structure of a glucocerebrosidase molecule or a portion thereof, wherein the model represents the glucocerebrosidase molecule or the portion thereof at a resolution of 2.9 angstroms or higher, and/or wherein the glucocerebrosidase molecule is partially glycosylated.

As described hereinabove, such a computer generated model can be used for facilitating computational identification of an optimal Gaucher disease drug.

As described hereinabove, the method of identifying a compound of the present invention comprises using a set of structure coordinates defining a 3D structure of a normal glucocerebrosidase molecule of the present invention, or a portion thereof, for computationally generating a set of structure coordinates defining a predicted 3D structure of a mutant glucocerebrosidase molecule, or a portion thereof. As described hereinabove, the method of identifying the compound further comprises using a set of structure coordinates defining a 3D structure of the mutant glucocerebrosidase molecule, or portion thereof, and optionally a set of structure coordinates defining a 3D structure of the mutant glucocerebrosidase molecule, or portion thereof, for computationally identifying a compound capable of interacting with the mutant glucocerebrosidase molecule in such a way as to correct the impaired enzymatic activity thereof.

Using a set of structure coordinates defining an experimentally determined 3D structure of a molecule, such as a normal glucocerebrosidase molecule of the present invention, for generating a set of structure coordinates defining a predicted 3D structure of an altered form of such a molecule, such as a mutant glucocerebrosidase molecule of the present invention relative to the normal glucocerebrosidase molecule may be advantageously effected using widely available software applications and widely practiced computational modeling techniques. Ample guidance regarding the selection and exploitation of such computer programs is provided in the literature of

the art (for example, refer to Mosyak L. *et al.*, 1995. Nat Struct Biol. 2:537-47). In particular, the set of structure coordinates defining the predicted 3D structure of the mutant glucocerebrosidase molecule or portion thereof may be generated using a program capable of predicting the effect of an amino acid substitution on the 3D structure of a protein whose experimentally determined 3D structure is known. Preferably, generating such a set of structure coordinates is effected according to the method described in Example 2 of the Examples section which follows.

Examples of suitable programs for predicting the 3D structure of a molecule include MODELLER (Marti-Renom M.A. *et al.*, 2000. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325), ESYPRED3D (Lambert C. *et al.*, 2002. Bioinformatics 18, 1250-6), and SWISS-MODEL (Guex N. and Peitsch MC., 1997. Electrophoresis 18, 2714-2723).

Computational identification of the compound of the present invention may be achieved by the ordinarily skilled artisan using the computational methods and software described below. Such computational identification of a compound having a desired biochemical effect on biomolecule may often be referred to in the art as "rational drug design" or "computational drug design".

Computational drug design:

Computational drug design is a potent means of identifying molecules capable of regulating enzyme activity which, for example, has notably been used to identify molecules capable of therapeutically regulating pathogenic HIV protease (Lam *et al.*, 1994. Science 263, 380; Wlodawer *et al.*, 1993. Ann Rev Biochem. 62, 543; Appelt, 1993. Perspectives in Drug Discovery and Design 1, 23; Erickson, 1993. Perspectives in Drug Discovery and Design 1, 109), and bcr-abl tyrosine kinase enzymatic activity (Mauro MJ. *et al.*, 2002. J Clin Oncol. 20, 325-34), and thereby to provide effective pharmacological cures for human acquired immunodeficiency syndrome (AIDS) caused by human immunodeficiency virus (HIV)), and a human cancer (chronic myeloid leukemia), respectively.

As described hereinabove, methods of computationally identifying the compound of the present invention may advantageously comprise screening a chemical structure database ("3D database"). This may be effectively performed using software employing "scanner" type algorithms which employ a set of structure coordinates defining a 3D structure of a mutant glucocerebrosidase molecule of the

present invention or a portion thereof, and of a chemical structure of a candidate compound stored in the database to computationally model the "docking" of the screened compound structure with the mutant glucocerebrosidase molecule or portion thereof, and to model the resultant structure of the mutant glucocerebrosidase molecule or portion *per se* of that of a complex thereof with the screened compound following association of the mutant glucocerebrosidase molecule or portion thereof with the screened compound. Iterating this process with each of a plurality of chemical structures stored in the database therefore enables computational screening of such a plurality to identify a compound having a structure enabling correction of the defective structure of the mutant glucocerebrosidase molecule or portion thereof, and hence to identify a compound potentially correcting the impaired enzymatic activity of the mutant glucocerebrosidase molecule.

Examples of suitable chemical structure databases for identifying the compound of the present invention include ISIS (MDL Information Systems, San Leandro, http://www.molinfo.com), MACCS-3D (Martin, Y. C., 1992. J. Med. Chem. 35, 2145-2154), The Cambridge Structural Database (CSD; http://www.ccdc.cam.ac.uk/prods/csd/csd.html), Fine Chemical Database (reviewed in Rusinko A., 1993. Chem Des Auto. News 8, 44-47), and the NCBI's Molecular Modeling DataBase: MMDB; http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml.

To identify the compound of the present invention via *de novo* computational drug design, or via modification of a known chemical structure, software comprising "builder" type algorithms utilizes a set of structure coordinates defining a 3D structure of a portion of the mutant glucocerebrosidase molecule, preferably a mutant portion thereof, and the 3D structures of basic chemical building blocks to computationally assemble a compound of the present invention. Such an approach may be employed for structurally refining a compound of the present invention identified, for example, via chemical database screening as described above.

Ample guidance for computationally identifying a compound having a desired effect on an enzyme via software employing such "scanner" and "builder" type algorithms is available in the literature of the art (for example, refer to: Halperin I. *et al.*, 2002. Proteins 47, 409-43; Gohlke H. and Klebe G., 2001. Curr Opin Struct Biol. 11, 231-5; Zeng J., 2000. Comb Chem High Throughput Screen. 3, 355-62; and

RACHEL: Theory of drug design, http://www.newdrugdesign.com/Rachel_Theory.htm#Software), and described in further detail hereinbelow.

Criteria employed by software programs used in computational drug design for qualifying the binding of screened compound structures with target portions of enzymes include gap space, hydrogen bonding, electrostatic interactions, van der Waals forces, hydrophilicity/hydrophobicity, etc. Generally, the greater the contact area between the screened chemical structure and the mutant glucocerebrosidase molecule or portion thereof, the lower the steric hindrance, the lower the "gap space", the greater the number of hydrogen bonds, and the greater the sum total of the van der Waals forces between the screened molecule and the mutant portion of the mutant glucocerebrosidase molecule, the greater will be the capacity of the screened molecule to bind with the mutant portion of the glucocerebrosidase molecule. The "gap space" refers to unoccupied space between the van der Waals surface of a screened molecule positioned within a binding pocket and the surface of the binding pocket defined by amino acid residues in the binding pocket. Gap space may be identified, for example, using an algorithm based on a series of cubic grids surrounding the docked chemical structure, with a user-defined grid spacing, and represents volume that could advantageously be occupied by a modifying the docked chemical structure positioned in contact with the mutant portion of the glucocerebrosidase molecule.

Contact area between screened compound structures and the mutant glucocerebrosidase molecule or portion thereof may be directly calculated from the coordinates of the compounds in docked conformation using the MS program (Connolly ML., 1983. Science 221, 709-713).

Suitable software employing "scanner" type algorithms include, for example, docking software such as GRAM, DOCK, or AUTODOCK (reviewed in Dunbrack *et al.*, 1997. Folding and Design 2, 27), AFFINITY software of the INSIGHTII package (Molecular Simulations Inc., 1996, San Diego, Calif.), GRID (Goodford PJ., 1985. "A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules", J. Med. Chem. 28, 849-857; GRID is available from Oxford University, Oxford, UK), and MCSS (Miranker A. and Karplus M., 1991. "Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method", Proteins: Structure Function and Genetics 11, 29-34; MCSS is available from Molecular Simulations, Burlington, Mass.).

The AUTODOCK program (Goodsell DS. and Olson AJ., 1990. Proteins: Struct Funct Genet. 8, 195-202; available from Scripps Research Institute, La Jolla, Calif.) helps in docking screened compound structures to target portions of target biomolecules in a flexible manner using a Monte Carlo simulated annealing approach. The procedure enables a search without bias introduced by the researcher. This bias can influence orientation and conformation of a screened molecule in the targeted binding pocket.

The DOCK program (Kuntz ID. *et al.*, 1982. J Mol Biol. 161, 269-288; available from University of California, San Francisco), is based on a description of the negative image of a space-filling representation of the target portion of the target biomolecule, and includes a force field for energy evaluation, limited conformational flexibility and consideration of hydrophobicity in the energy evaluation.

Modeling or docking may be followed by energy minimization with standard molecular mechanics force fields or dynamics with programs such as CHARMM (Brooks BR. *et al.*, 1983. J Comp Chem. 4, 187-217) or AMBER (Weiner SJ. *et al.*, 1984. J Am Chem Soc. 106, 765-784).

As used herein, "minimization of energy" means achieving an atomic geometry of a chemical structure via systematic alteration such that any further minor perturbation of the atomic geometry would cause the total energy of the system as measured by a molecular mechanics force-field to increase. Minimization and molecular mechanics force fields are well understood in computational chemistry (for example, refer to Burkert U. and Allinger NL., "Molecular Mechanics", ACS Monograph 177, pp. 59-78, American Chemical Society, Washington, D.C. (1982)).

Programs employing "builder" type algorithms include LEGEND (Nishibata Y. and Itai A., 1991. Tetrahedron 47, 8985; available from Molecular Simulations, Burlington, Mass.), LEAPFROG (Tripos Associates, St. Louis, Mo.), CAVEAT (Bartlett, PA. *et al.*, 1989. Special Pub Royal Chem Soc. 78, 182-196; available from University of California, Berkeley), HOOK (Molecular Simulations, Burlington, Mass.), and LUDI (Bohm HJ., 1992. J. Comp Aid Molec Design 6, 61-78; available from Biosym Technologies, San Diego, Calif.).

The CAVEAT program suggests binding structures of screened compound based on desired bond vectors. The HOOK program proposes docking sites by using multiple copies of functional groups in simultaneous searches. LUDI is a program

based on fragments rather than on descriptors which proposes somewhat larger fragments to match with a target binding structure and scores its hits based on geometric criteria taken from the Cambridge Structural Database (CSD), the Protein Data Bank (PDB) and on criteria based on binding data. LUDI may be advantageously employed to calculate the inhibition constant of a docked structure. Activation constants (Ka values; opposite of inhibition constant, Ki) of structures of compounds in the final docking positions can be evaluated using LUDI software.

During or following computational selection of a screened structure of a compound, docking of an intermediate chemical structure or of the structure of the screened compound with the mutant glucocerebrosidase molecule or portion thereof may be visualized via structural models, such as 3D models, thereof displayed on a computer screen, so as to advantageously allow user intervention during the rational drug design to optimize a chemical structure.

Software programs useful for displaying such 3D structural models, include RIBBONS (Carson, M., 1997. Methods in Enzymology 277, 25), O (Jones, TA. *et al.*, 1991. Acta Crystallogr. A47, 110), DINO (DINO: Visualizing Structural Biology (2001) http://www.dino3d.org); and QUANTA, INSIGHT, SYBYL, MACROMODE, ICM, MOLMOL, RASMOL and GRASP (reviewed in Kraulis, J., 1991. Appl Crystallogr. 24, 946).

Other molecular modeling techniques may also be employed in accordance with this invention (for example, refer to: Cohen NC. et al, 1990. "Molecular Modeling Software and Methods for Medicinal Chemistry", J. Med. Chem. 33, :883-894; Navia M. A. and Murcko M. A., 1992. "The Use of Structural Information in Drug Design", Current Opinions in Structural Biology 2, 202-210). For example, where the structures of test compounds are known, a model of the test compound may be superimposed over the model of the structure of the invention. Numerous methods and techniques are known in the art for performing this step, any of which may be used (for example, refer to: Farmer P. S., "Drug Design", Ariens EJ. (ed.), Vol. 10, pp 119-143 (Academic Press, New York, 1980); U.S. Pat. No. 5,331,573; U.S. Pat. No. 5,500,807; Verlinde C., 1994. Structure 2, 577-587; and Kuntz ID., 1992. Science 257, 1078-108).

Thus, using such computational methods a large number of 3D structures of screened compounds may be quickly and easily examined and expensive and lengthy

biochemical testing avoided. Moreover, the need for actual synthesis of many compounds is effectively eliminated.

Various aspects of the present invention may be advantageously practiced using a computing platform **30** (Figure 4) which generates a graphic output of a model of a glucocerebrosidase molecule of the present invention or a portion thereof via display **32**. The computing platform generates graphic representations of the model via processing unit **34** which processes a set of structure coordinates defining the 3D structure in a retrievable format in data storage device **36**. Examples of computer readable media which can be used for storing a set of structure coordinates include conventional computer hard drives, floppy disks, DAT tape, CD-ROM, and other magnetic, magneto-optical, optical, floptical, and other media which may be adapted for use with computing platform **30**.

Preferably the set of structure coordinates of the present invention is in PDB format for convenient processing by such software applications. Most or all of these software applications, and others as well, are downloadable from the World Wide Web.

Those of ordinary skill in the art will appreciate that a set of structure coordinates defining a 3D structure of a molecule is a relative set of points that define a shape in three dimensions. Thus, it is possible that a different set of coordinates, for example a set of coordinates utilizing a different frame of reference and/or different units, could define a similar or identical shape. Moreover, it will be understood that slight variations in the individual coordinates will have little effect on overall shape. Such variations in coordinates may result, for example, from mathematical manipulations of the coordinates. For example, coordinates can be manipulated by crystallographic permutations of the atomic coordinates, fractionalization of the coordinates, integer additions or subtractions to sets of the coordinates, inversion of the coordinates or any combination of the above. Alternatively, modifications in a crystal structure from which the coordinates are derived due to mutations, additions, substitutions, or other changes in any of the components that make up the crystal could also account for variations in the coordinates. If such variations are within an acceptable standard error as compared to the original coordinates, the resulting 3D shape is considered to be the same.

Thus, the computational methods, software programs, and algorithms

described hereinabove can be used by the ordinarily skilled practitioner for efficiently computationally identifying the compound of the present invention.

Once the compound of the present invention is computationally identified it may be ordered from a commercial chemical library such as, for example, one held by a large chemical company such as Merck, Glaxo Welcome, Bristol Meyers Squib, Monsanto/Searle, Eli Lilly, Novartis, Pharmacia UpJohn, and the like. The compound of the present invention may also be ordered via the World Wide Web (Internet) via companies such as Chemcyclopedia (http://www.mediabrains.com/client/chemcyclop /BG1/search.asp). Alternatively, the compound of the present invention may be synthesized *de novo* using standard chemical and/or biological synthesis techniques, as appropriate to the molecular type. Ample guidance for synthesis of molecules typically identified via the above-described rational drug design methodology is provided in the literature of the art. For biological synthesis of molecules, such as polypeptides and nucleic acids, refer, for example to: Sambrook *et al.*, infra; and associated references in the Examples section which follows. For guidance regarding chemical synthesis of molecules, refer, for example to the extensive guidelines provided by The American Chemical Society (http://www.chemistry.org/ portal/Chemistry). One of ordinary skill in the art, such as, for example, a chemist, will possess the required expertise for chemical synthesis of molecules such as the compound of the present invention.

As described hereinabove, following computational identification of the compound of the present invention, the capacity of the compound to correct the impaired enzymatic activity of the mutant glucocerebrosidase molecule is preferably biochemically qualified. Such biochemical qualification may be achieved as described hereinabove.

Thus, the present invention provides a crystallized glucocerebrosidase molecule having a unique X-ray diffraction capacity enabling for the first time generation of structure coordinates defining at atomic resolution an essentially complete 3D structure of a glucocerebrosidase molecule. Such structure coordinates of the present invention enabled for the first time generation of a model representing such a structure. Furthermore, such a computing platform of the present invention enabled for the first time generation of a set of structure coordinates defining at atomic resolution an essentially complete, optimally accurate, predicted 3D structure

of a mutant glucocerebrosidase molecule associated with Gaucher disease. Moreover, such computer generated models of the present invention enable for the first time identification of a compound capable of optimally correcting an impaired enzymatic activity of a mutant glucocerebrosidase molecule associated with Gaucher disease, and being characterized by essentially any desired biological, chemical and/or physical characteristics. Hence, the present invention enables identification of optimal Gaucher disease drugs targeted towards essentially any mutant human glucocerebrosidase molecule of interest.

Additional objects, advantages, and novel features of the present invention will become apparent to one ordinarily skilled in the art upon examination of the following examples, which are not intended to be limiting. Additionally, each of the various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below finds experimental support in the following examples.

EXAMPLES

Reference is now made to the following examples, which together with the above descriptions, illustrate the invention in a non limiting fashion.

Generally, the nomenclature used herein and the laboratory procedures utilized in the present invention include molecular, biochemical, microbiological and recombinant DNA techniques. Such techniques are thoroughly explained in the literature. See, for example, "Molecular Cloning: A laboratory Manual" Sambrook *et al.*, (1989); "Current Protocols in Molecular Biology" Volumes I–III Ausubel, R. M., ed. (1994); Ausubel *et al.*, "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Maryland (1989); Perbal, "A Practical Guide to Molecular Cloning", John Wiley & Sons, New York (1988); Watson *et al.*, "Recombinant DNA", Scientific American Books, New York; Birren *et al.* (eds) "Genome Analysis: A Laboratory Manual Series", Vols. 1–4, Cold Spring Harbor Laboratory Press, New York (1998); methodologies as set forth in U.S. Pat. Nos. 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057; "Cell Biology: A Laboratory Handbook", Volumes I–III Cellis, J. E., ed. (1994); "Current Protocols in Immunology" Volumes I–III Coligan J. E., ed. (1994); Stites *et al.* (eds), "Basic and Clinical Immunology"

(8th Edition), Appleton & Lange, Norwalk, CT (1994); Mishell and Shiigi (eds), "Selected Methods in Cellular Immunology", W. H. Freeman and Co., New York (1980); available immunoassays are extensively described in the patent and scientific literature, see, for example, U.S. Pat. Nos. 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; 4,098,876; 4,879,219; 5,011,771 and 5,281,521; "Oligonucleotide Synthesis" Gait, M. J., ed. (1984); "Nucleic Acid Hybridization" Hames, B. D., and Higgins S. J., eds. (1985); "Transcription and Translation" Hames, B. D., and Higgins S. J., eds. (1984); "Animal Cell Culture" Freshney, R. I., ed. (1986); "Immobilized Cells and Enzymes" IRL Press, (1986); "A Practical Guide to Molecular Cloning" Perbal, B., (1984) and "Methods in Enzymology" Vol. 1–317, Academic Press; "PCR Protocols: A Guide To Methods And Applications", Academic Press, San Diego, CA (1990); Marshak *et al.*, "Strategies for Protein Purification and Characterization – A Laboratory Course Manual" CSHL Press (1996); all of which are incorporated by reference as if fully set forth herein. Other general references are provided throughout this document. The procedures therein are believed to be well known in the art and are provided for the convenience of the reader.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below.

EXAMPLE 1

The 3D atomic structure of human glucocerebrosidase

Gaucher disease is a highly debilitating disease caused by mutant glucocerebrosidase molecules having impaired enzymatic activity, and for which no satisfactory treatment is available. An optimal drug for treatment of Gaucher disease would be of optimally small dimensions, have a non-polypeptidic composition, and would be capable of correcting impaired glucocerebrosidase enzymatic activity in mutant glucocerebrosidase molecules associated with pathogenesis of the disease. Ideally, such compounds could be computationally identified using computing

platforms capable of generating atomic resolution models representing 3D atomic structures of glucocerebrosidase molecules. The prior art, however, has failed to provide computing platforms suitable for performing such rational drug design. As described below, while reducing the present invention to practice, the present inventors have overcome the limitations of the prior art by producing a computing platform capable of generating a model representing at atomic resolution the essentially complete, experimentally determined 3D structure of human glucocerebrosidase, thereby enabling computational identification of optimal Gaucher disease drugs.

Materials and Methods:

Crystallization: A 5 mg aliquot of Cerezyme®, a recombinant variant of human glucocerebrosidase having the natural Arg495 residue substituted with a His residue, and oligosaccharide chains of glycosylated N-linked glycosylation sites modified to terminate in mannose sugars, was dialyzed overnight against phosphate-buffered saline (PBS) pH 7.0 and deglycosylated by incubation at 25 °C for 88 hours with 150 units N-glycosidase F. Cerezyme®, which exhibits an enzymatic activity identical to that of natural human glucocerebrosidase (Grabowski GA. *et al.*, 1995. Ann Intern Med. 122:33-9) and is thereby properly folded, was employed due to its availability as the active ingredient of the commercial drug Cerezyme® (Genzyme Corporation, Cambridge, MA, USA). Deglycosylation was monitored by the reduction in molecular weight of the enzyme by SDS-PAGE, and mass spectrometry revealed removal of 7-14 sugar residues. The concentration of the deglycosylated enzyme was adjusted to 10 mg/ml in 1 millimolar 2-morpholinoethanesulfonic acid (MES) pH 6.6, 0.1 molar NaCl , 0.02 % NaN_3 using a Centricon YM-10 centrifugal concentrator equipped with a molecular weight cut-off filter of about 10 kDa. Crystals were obtained via the hanging-drop technique at room temperature using a drop composed of 1.5 microliters of the 10 mg/ml glucocerebrosidase solution and 1.5 microliters of mother liquor (1 molar $(NH_2)_2SO_4$ pH 4.6, 0.17 molar guanidine hydrochloride, 0.02 molar KCl, 0.1 molar sodium acetate buffer, pH 4.6. Crystals were cryo-protected with a gradient of 5-25 % glycerol. A heavy-atom derivative was obtained by soaking for 3 days in a 1:125,000 dilution of $KHgI_2$ in mother liquor.

Data collection: X-ray data were collected at 100K at 3 wavelengths around the Hg LIII-edge on beamline ID14-4, and a native data set on beamline BM14 at the

European Synchrotron Radiation Facility (Grenoble, France). The enzyme crystallized in a $C222_1$ space group with two molecules in the asymmetric unit. The data was processed using MOSFLM/SCALA (Leslie AGW., 1992. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography 26) and DENZO/SCALEPACK (Otwinowski Z. and Minor W., 1997. Methods Enzymol. 276:307-326) software. The space group and cell dimensions obtained were similar to those recently reported for crystals of the glycosylated form of the enzyme which diffracted, however, to significantly lower resolution (Roeber D. *et al.*, 2003. Acta Cryst. D59:343-344). Data collection statistics are shown in Table 1.

Table 1. Data collection statistics

	Hg inflection	Hg peak	Hg remote	Native
Wavelength (angstroms)	1.0092	1.0075	1.0015	0.8856
Unit cell (angstroms)	109.2, 286.1, 91.5	109.2, 286.7, 91.5	108.9, 284.1, 91.2	107.7, 285.2, 91.8
Resolution range (angstrom)	26-2.35	26-2.27	26-2.27	10-2.00
No. of unique reflections	58,766	64,819	64,311	93,248
Completeness (%)	97.1 (98.3)*	96.1 (87.8)*	97.5(99.1)*	98.4 (98.3)*
I/σ(I)	9.6 (2.6)*	9.52 (2.1)*	9.1 (2.3)*	7.4 (1.6)*
Rsym(I) (%)	7.0 (24.7)*	7.2 (29.0)*	7.7 (29.4)*	8.4 (37.3)*

*Data for outer shell are in parentheses.

Structure determination and refinement: Three Hg sites were located based on their anomalous difference using SHELXD software (Uson I. and Sheldrick GM., 1999. Curr Opin Struct Biol. 9:643-648). The Hg sites were refined, and experimental phases to 2.3 angstroms were calculated from the MAD data using SHARP software (Fortelle E. and Bricogne G., 1997. Methods Enzymol. 276:472-494), resulting in an overall FOM of 0.403. Phases were improved by applying solvent flipping density modification using SOLOMON software (Abrahams JP. and Leslie AG., 1996. Acta Crystallogr D52:30-42), resulting in an overall FOM of 0.851. Automated tracing procedure performed using ARP/wARP software (Perrakis A. *et al.*, 1999. Nature Struct Biol. 6:458-463), using native amplitudes to 2.0 angstroms, coupled to the experimental phases, resulted in tracing of about 95 % of the two polypeptide chains. The σA map shows all 497 residues in both molecules. Final tracing was performed manually using O software (Jones TA. *et al.*, 1991. Acta Cryst A47:110-119). Refinement of the two molecules was performed using REFMAC (Murshudov GN. *et*

al., 1999. Acta Cryst D55:247-255) and CNS (Brunger AT. *et al.*, 1998. Acta Cryst. D54:905-921) software at 2.0 angstroms, with an overall rmsd of 0.29 angstroms for Cα atoms between the two molecules. The maps showed a single glycosylation site at Asn19 with one N-acetylglucosamine moiety on one molecule and two on the other. 809 water molecules and 15 sulfate ions were assigned. Refinement and model statistics are shown in Table 2.

Table 2. Refinement and model statistics.

Resolution range (angstroms)	14.4-2.0
No. of reflections	88,501
R-factor (%) work, free	19.5, 23.0
No. of atoms:	
Protein (994 residues)	7,859
Hetero (carbohydrate, solvent)	1,056
Average B-factors (square angstroms)	28.4
RMSD from ideal values:	
Bond length (angstroms)	0.005
Bond angle (degrees)	1.3
Dihedral angles (degrees)	23.8
Improper torsion angles (degrees)	0.86
Estimated coordinate error:	
Low resolution cutoff (angstroms)	5.0
ESD from Luzzati plot (angstroms)	0.23
ESD from sigmaA (angstroms)	0.32
Ramachandran outliers (%)	3.1

Experimental Results:

Following extensive empirical experimentation, highly purified, highly ordered human glucocerebrosidase polypeptide crystals were successfully grown. X-ray crystallographic analysis of such crystals generated Protein DataBank (PDB) a set of structure coordinates defining the essentially complete 3D atomic structure of human glucocerebrosidase polypeptide at 2.0 angstroms (Table 4, enclosed CD-ROM). The refined structure (R-factor 19.9 %, R-free 23.2 %) indicated that crystallized glucocerebrosidase contains two enzyme molecules per asymmetric unit. The overall fold of the enzyme comprises 3 domains (Figures 1b and 1c). Domain I (residues 1-27 and 383-414) consists of one major 3-stranded anti-parallel beta-sheet flanked by a perpendicular N-terminal strand and a loop. It contains two disulfide bridges (residues 4-16 and 18-23), which may be required for correct folding [Beutler E. and Grabowski GA., in: "The Metabolic and Molecular Bases of Inherited

Disease", Scriver CR. *et al.* (eds.), McGraw-Hill Inc., pp. 3635-3668 (2001)]. Glycosylation, which is essential for catalytic activity *in-vivo* (Berg-Fussman A. *et al.*, 1993. J Biol Chem. 268:14861-14866), is seen in the crystal structure at N19. Domain II (residues 30-75, 431-497) consists of two closely-associated beta-sheets forming an independent domain resembling an immunoglobulin (Ig) fold (Orengo *et al.*, 1997. Structure 5:1093; Westhead *et al.*, 1999. Protein Sci. 8:897). Domain III (residues 76-381, 416-430) is a (beta/α)$_8$ (TIM) barrel containing the catalytic site, consistent with homology to GH-A clan members (Fabrega S. *et al.*, 2002. J Soc Biol. 196:151-60; Henrissat and Bairoch, 1993. Biochem. J. 316:695-696). It contains three free cysteines (residues 126, 248 and 342). Domains II and III appear to be connected by a hinge region located between helix 8 and strand 1 of the TIM-barrel and the loops connecting the two domains, whereas domain 1 tightly interacts with domain III.

Site-directed mutagenesis and homology modeling of glucocerebrosidase (Fabrega S. *et al.*, 2000. Glycobiology 10:1217-24; Fabrega S. *et al.*, 2002. J Soc Biol. 196:151-60) suggested that E235 is the acid/base catalyst and E340 the nucleophile (Miao S. *et al.*, 1994. J Biol Chem. 269:10975-10978). These two residues are located near the C-termini of strands 4 and 7 (Figures 1b and 2a) in domain III, with an average distance between their carboxyl oxygens of 5.2 angstroms for the two glucocerebrosidase molecules in the structure, consistent with retention of the anomeric carbon upon cleavage, rather than inversion (Davies G. and Henrissat B., 1995. Structure 3:853-859). Thus residues D443 and D445, now known to be located in the Ig-like domain (Figure 1b), cannot be directly involved in catalysis even though they are covalently labeled (Dinur T. *et al.*, 1986. Proc Natl Acad Sci U S A. 83:1660-1664) by the irreversible glucocerebrosidase inhibitor, conduritol-B-epoxide (Legler G., 1977. Methods Enzymol. 46:368-381). Substrate docking shows that only the glucose moiety and adjacent glycoside bond of glucosylceramide fit within the active site pocket (Figure 2b), implying that the two glucosylceramide hydrocarbon chains either remain embedded in the lipid bilayer during catalysis or alternatively interact with saposin C. In addition, an annulus of hydrophobic residues surrounds the entrance to the active site (Figure 2b), and may facilitate interaction of glucocerebrosidase with the lysosomal membrane or with saposin C (Legler G., 1977. Methods Enzymol. 46:368-381).

Of the approximately 200 known glucocerebrosidase mutations (Figures 1d-e), many are rare and restricted to a few individuals. Most mutations partially or entirely decrease catalytic activity (Meivar-Levy, I. *et al.*, 1994. Biochem. J. 303:377-382) or are believed to reduce glucocerebrosidase stability (Grace ME. *et al.*, 1994. J Biol Chem. 269:2283-2291). The most common mutation, N370S, accounts for 70 % of the mutant alleles in Ashkenazi Jews and 25 % of non-Jewish patients [Beutler E. and Grabowski GA., in: "The Metabolic and Molecular Bases of Inherited Disease", Scriver CR. *et al.* (eds.), McGraw-Hill Inc., pp. 3635-3668 (2001)]. This mutation predisposes to Type 1 disease and always precludes neurological involvement, implying that it causes relatively minor changes in glucocerebrosidase structure, and hence catalytic activity. Consistent with this is the localization of N370 to the longest α-helix (helix 7) in glucocerebrosidase, which is located at the interface of domains II and III, but is too far from the active site to directly participate in catalysis. Interestingly, a number of other mutations are found in this helix, all of which appear to point into the TIM-barrel (Figure 3).

Seven aromatic side chains (F128, W179, Y244, F246, Y313, W381, F397) line one side of the active site pocket, which may be involved in substrate recognition and has been observed in other beta-glycosidases (Chi YI. *et al.*, 1999. FEBS Lett. 445:375-383; Henrissat B. and Bairoch A., 1993. Biochem J. 293:781-788). The common mutation V394L [Beutler E. and Grabowski GA., in: "The Metabolic and Molecular Bases of Inherited Disease", Scriver CR. *et al.* (eds.), McGraw-Hill Inc., pp. 3635-3668 (2001); Table 3] might perturb this lining as the bulkier leucine side chain could cause a conformational change in 2 members of the lining, Y244 and F246.

Several other mutations occur near the active site, i.e., H311, R341 and C342, and may directly affect catalytic activity. In contrast, two relatively common mutations (Table 3), R463C and R496H, which predispose to mild disease [Beutler E. and Grabowski GA., in: "The Metabolic and Molecular Bases of Inherited Disease", Scriver CR. *et al.* (eds.), McGraw-Hill Inc., pp. 3635-3668 (2001)], are located in the Ig-like domain at a considerable distance from the active site (Figure 1b).

Table 3. Most common single amino acid substitutions in glucocerebrosidase associated with Gaucher disease.

Mutation*	Phenotype	Features	Enzyme activity	Structural features
N370S	Mild	70 % of mutant alleles in Ashkenazi Jews; invariably predisposes to mild (type 1) disease	Reduced activity; stable protein	On longest helix (helix 7) in protein at interface of domains II and III. Several other mutations are found on this helix (see Figure 3)
V394L	Severe		Reduced activity; stable protein	Near aromatic residues that line one side of the active site pocket; may disrupt this lining and thus catalytic activity
D409H	Severe		Highly reduced activity; unstable protein	On domain I, suggesting regulatory or structural role for this domain
L444P	Severe	Most common mutation predisposing to severe (type 2/3) disease	Reduced activity; unstable protein	Hydrophobic core of Ig-like domain (domain II), which may lead to unstable protein due to disruption of the hydrophobic core and altered folding of this domain
R463C	Mild		Reduced activity; stable protein	On Ig-like domain, distant from the active site
R496H	Mild			On Ig-like domain, distant from the active site

* For references pertaining to each of these mutations, refer to Beutler E. and Grabowski GA., in: "The Metabolic and Molecular Bases of Inherited Disease", Scriver CR. et al. (eds.), McGraw-Hill Inc., pp. 3635-3668 (2001)]

Unexpectedly, L444, which is mutated relatively frequently to proline or arginine, and invariably predisposes to severe neuronopathic disease [Beutler E. and Grabowski GA., in: "The Metabolic and Molecular Bases of Inherited Disease", Scriver CR. *et al.* (eds.), McGraw-Hill Inc., pp. 3635-3668 (2001); Erikson, A. *et al.* in "Gaucher's Disease", Zimran, A. (ed.), Vol. 10, Bailliere Tindall, London, pp. 711-723 (1997)], is located in the hydrophobic core of the Ig-like domain (Figure 1b). Either of the two L444 mutations would likely cause a local conformational change by disrupting the hydrophobic core, resulting in altered folding of this domain (Morel N. *et al.*, 1999. Mol Pharmacol. 55:982-992), consistent with the assumption that these mutations produce unstable protein (Grace ME. *et al.*, 1994. J Biol Chem. 269:2283-2291). This implies an important regulatory or structural function for domain II, perhaps in interacting with saposin C and/or acidic phospholipids. Interestingly, beta-hexosaminidase and other family 20 glycosidases have a similar non-catalytic domain whose role is unknown (Mark BL. *et al.*, 2001. J Biol Chem. 276:10330-7).

The saposin C structure has recently been determined by NMR (PDB ID code 1M12), but its coordinates are on hold. However, the structure of its homolog saposin B (Ahn VE. *et al.*, 2003. Proc Natl Acad Sci U S A. 100:38-43), reveals that the putative active form is a dimer in which a large hydrophobic cavity sequesters the acyl chains of cerebroside sulfate, and may serve to present it appropriately for hydrolysis by arysulfatase A. It cannot yet be established whether such a mechanism would explain the role of saposin C as an activator of glucocerebrosidase since the limited sequence homology (less than 14 %) between saposins B and C does not permit accurate modeling of the latter. However, the Ig-like domain of glucocerebrosidase may regulate the interaction of glucocerebrosidase with either the lipid bilayer, with saposin C, or both. Finally, the lack of known viable mutations in residues 14-20 of domain I and in the connecting strand (1-10) and loop (21-27), with the exception of the conserved mutation V15L, together with the 7 known mutations in the C-terminal strand of this domain (401-414), including the common severe mutation D409H which results in unstable protein [Beutler E. and Grabowski GA., in: "The Metabolic and Molecular Bases of Inherited Disease", Scriver CR. *et al.* (eds.), McGraw-Hill Inc., pp. 3635-3668 (2001)], suggests that domain I also has an important regulatory or structural role.

In summary, the catalytic domain of glucocerebrosidase consists of a $(\beta/\alpha)_8$ (TIM) barrel found also in the GH-A glycosidase clan http://afmb.cnrs-mrs.fr/CAZY/GH_30.html. The catalytic residue E235 is H-bonded to H311 and Y313, and E340 is H-bonded to R120. The catalytic residues, E340 and E235, are hydrogen-bonded to R120 and to Y313 and H311, respectively, with the distance between the two glutamates consistent with a catalytic mechanism of retention. N370 is located on helix 7, which is at the interface of the TIM-barrel and a separate immunoglobulin-like domain on which L444 is located, implying a key regulatory or structural role for this non-catalytic domain.

Conclusion: The presently described crystallization method enabled for the first time relative to the prior art generation of crystals of human glucocerebrosidase having an X-ray diffraction capacity enabling generation of sets of structure coordinates defining at atomic resolution the essentially complete 3D structure of human glucocerebrosidase. Such sets of structure coordinates were used for the first time for generating computer-generated models representing at an essentially

complete 3D atomic structure of human glucocerebrosidase. As such, the presently described computing platforms enable computational identification of compounds optimally capable of correcting impaired enzymatic activity of mutant glucocerebrosidase molecules associated with pathogenesis of Gaucher disease, and thereby identification of Gaucher disease drugs overcoming all shortcomings of prior art Gaucher disease drugs.

EXAMPLE 2

The essentially complete, optimally accurate, predicted 3D atomic structure of human glucocerebrosidase mutants associated with Gaucher disease

As described above, whole enzyme replacement therapy using recombinant glucocerebrosidase is currently the treatment of choice for treatment of Gaucher disease. Such treatment, however, suffers from numerous significant drawbacks, as described above. Hence, novel and improved Gaucher disease drugs are urgently required. Optimal Gaucher disease drugs, would be compounds having optimally small dimensions and a non-polypeptidic chemical composition, and would be capable of interacting with mutant glucocerebrosidase molecules associated with Gaucher disease in such a way as to correct impaired enzymatic activity thereof. Ideally, such compounds could be computationally identified using computing platforms capable of generating essentially complete, experimentally determined, models representing 3D atomic structures of human glucocerebrosidase molecules. The prior art has, however, failed to provide such computing platforms. While reducing the present invention to practice, the present inventors have provided computing platforms overcoming such prior art limitations, as follows.

Materials and Methods:

Set of PDB structure coordinates defining the essentially complete predicted structure of common human glucocerebrosidase mutants: Sets of structure coordinates defining the predicted, essentially complete, 3D structure of human glucocerebrosidase mutants Asn370Ser, Val394Leu, Asp409His, Leu444Pro, Arg463Cys, and Arg496His were generated using the set of structure coordinates corresponding to atom coordinates 1–3929 set forth in Table 4 (enclosed CD-ROM) which define an essentially complete, experimentally determined, 3D structure of normal human glucocerebrosidase, as described in Example 1 of the Examples section

above. Briefly, structure coordinates were generated using SHELXD software to locate the positions of the heavy atoms, SHARP and SOLOMON to get an initial electron density map, and CNS and O to refine and fit the chain into the map.

A set of structure coordinates defining the predicted structure of the set of amino acid residues predicted to have at least one atom positioned within 10 angstroms of at least one atom of a mutated amino acid were also identified for each of the aforementioned human glucocerebrosidase mutants associated with Gaucher disease. The structure of such a set of amino acid residues located within 10 angstroms of the mutated amino acid residue is hereinafter referred to as "10-angstrom radius structure". A similar 10-angstrom radius structure was defined for each mutation within the structure of the non-mutated enzyme. These non-mutated 10-angstrom radius structures are defined by sets of Table 4 (enclosed CD-ROM) coordinates defining the structures of sets of amino acid residues from the amino acid sequence of the non-mutated glucocerebrosidase molecule located at the same positions within the amino acid sequence of the non-mutated glucocerebrosidase molecule (SEQ ID NO: 1) as the amino acid residues included in the 10-angstrom radius structures of the corresponding mutant glucocerebrosidase molecule within the amino acid sequence of the mutated glucocerebrosidase molecule.

Results:

Set of structure coordinates defining the predicted, essentially complete, 3D structures of human glucocerebrosidase (Cerezyme®) mutants Asn370Ser (SEQ ID NO: 2), Val394Leu (SEQ ID NO: 3), Asp409His (SEQ ID NO: 4), Leu444Pro (SEQ ID NO: 5), Arg463Cys (SEQ ID NO: 6), and Arg496His (SEQ ID NO: 7) at 2.0 angstrom resolution were generated (Tables 11, 13, 15, 17, 19, and 21, respectively, enclosed CD-ROM). These structure coordinates were used to generate:

(i) a set of structure coordinates defining the 10-angstrom radius structure defined in relation to Ser370 in human glucocerebrosidase mutant Asn370Ser (Table 12, enclosed CD-ROM);

(ii) a set of structure coordinates defining the 10-angstrom radius structure defined in relation to Leu394 in human glucocerebrosidase mutant Val394Leu (Table 14, enclosed CD-ROM);

(iii) a set of structure coordinates defining the 10-angstrom radius structure defined in relation to His409 in human glucocerebrosidase mutant Asp409His (Table

16, enclosed CD-ROM);

(iv) a set of structure coordinates defining the10-angstrom radius structure defined in relation to Pro444 in human glucocerebrosidase mutant Leu444Pro (Table 18, enclosed CD-ROM);

(v) a set of structure coordinates defining the 10-angstrom radius structure defined in relation to Cys463 in human glucocerebrosidase mutant Arg463Cys (Table 20, enclosed CD-ROM); and

(vi) a set of structure coordinates defining the10-angstrom radius structure defined in relation to His496 in human glucocerebrosidase mutant Arg496His (Table 22, enclosed CD-ROM).

The amino acid residues forming the 10-angstrom radius structures of the human glucocerebrosidase mutants were identified as having the following amino acid residue coordinates according to the amino acid sequences of the mutant polypeptides:

(i) 76, 81, 285, 312, 314, 320, 324, 325, 336, 364–378, 423, and 433 for human glucocerebrosidase mutant Asn370Ser (SEQ ID NO: 2);

(ii) 244–247, and 390–397 for human glucocerebrosidase mutant Val394Leu (SEQ ID NO: 3);

(iii) 20, 21, 95–100, and 404–411 for human glucocerebrosidase mutant Asp409His (SEQ ID NO: 4);

(iv) 65–67, 440–447, 460–464, 468, and 469 for human glucocerebrosidase mutant Leu444Pro (SEQ ID NO: 5);

(v) 360–366, 443–446, 460–467, and 484–89 for human glucocerebrosidase mutant; Arg463Cys (SEQ ID NO: 6); and

(vi) 33–35, 69, 71, 450–456, 474–478, and 493–497 for human glucocerebrosidase mutant Arg496His (SEQ ID NO: 7).

The relevant above-described structure coordinates were used to computationally generate models representing the predicted, essentially complete, structure of human glucocerebrosidase mutants Asn370Ser, Val394Leu, Asp409His, Leu444Pro, Arg463Cys, and Arg496His at 2.0 angstrom resolution.

The relevant above-described sets of structure coordinates were used to computationally generate models representing the predicted 10-angstrom radius structure defined in relation to mutant amino acid residues Ser370, Leu394, His409,

Pro444, Cys463, and His496 in human glucocerebrosidase mutants Asn370Ser, Val394Leu, Asp409His, Leu444Pro, Arg463Cys, and Arg496His, respectively.

For comparison of the mutant 10-angstrom radius structures with corresponding structures of normal human glucocerebrosidase, the coordinates defining the structure of amino acid residues of the normal human glucocerebrosidase [(SEQ ID NO: 1), whose structure is defined by atom coordinates 1–3929 of Table 4 (enclosed CD-ROM)] having the same amino acid sequence position coordinates in the normal glucocerebrosidase polypeptide as those of the amino acid residues forming the mutant 10-angstrom radius structure in the mutant polypeptide were employed. Hence, the structures of the portions of the normal glucocerebrosidase corresponding to the 10-angstrom radius structures of the human glucocerebrosidase mutants Asn370Ser, Val394Leu, Asp409His, Leu444Pro, Arg463Cys, and Arg496His are defined by the sets of structure coordinates set forth in Tables 5-10 (enclosed CD-ROM), respectively.

Conclusion: The presently disclosed method enables for the first time generation of sets of structure coordinates defining with optimal accuracy essentially complete predicted 3D atomic structures of mutant human glucocerebrosidase molecules associated with Gaucher disease, in particular those of the most prevalent mutants Asn370Ser (SEQ ID NO: 2), Val394Leu (SEQ ID NO: 3), Asp409His (SEQ ID NO: 4), Leu444Pro (SEQ ID NO: 5), Arg463Cys (SEQ ID NO: 6), and Arg496His (SEQ ID NO: 7). The presently-described sets of structure coordinates were used for generating for the first time essentially complete optimally accurate computer-generated models representing such 3D structures. Hence, by virtue of such capacity, the presently described computing platforms enable for the first time computational identification of compounds capable of correcting *in-vivo*, with optimal efficacy and safety, the impaired enzymatic activity of essentially any given mutant glucocerebrosidase molecule associated with Gaucher disease. Hence, the presently disclosed method enables identification of optimal Gaucher disease drugs.

It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be

provided separately or in any suitable subcombination.

Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims. All publications, patents, patent applications, and sequences identified by their accession numbers mentioned in this specification are herein incorporated in their entirety by reference into the specification, to the same extent as if each individual publication, patent, patent application or sequence identified by its accession number was specifically and individually indicated to be incorporated herein by reference. In addition, citation or identification of any reference in this application shall not be construed as an admission that such reference is available as prior art to the present invention.

CD-ROM Content

The following CD-ROM is attached herewith:

Information provided as: File name / byte size / date of creation / operating system / machine format

CD-ROM1 (19 files):

1. Table 4.txt / 696,500 bytes / May 25, 2003 / Microsoft Windows Word 2000 / PC.
2. Table 5.txt / 15,562 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.
3. Table 6.txt / 7,676 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.
4. Table 7.txt / 8,220 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.
5. Table 8.txt / 9,920 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.
6. Table 9.txt / 13,252 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.
7. Table 10.txt / 12,288 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.
8. Table 11.txt / 283,630 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.
9. Table 12.txt / 16,156 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.
10. Table 13.txt / 283,846 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.
11. Table 14.txt / 8,168 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.
12. Table 15.txt / 283,918 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.
13. Table 16.txt / 8,672 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.
14. Table 17.txt / 283,702 bytes / May 25, 2003/ Microsoft Windows Word

2000 / PC.

15. Table 18.txt / 10,256 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.

16. Table 19.txt / 283,414 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.

17. Table 20.txt / 13,568 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.

18. Table 21.txt / 283,702 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.

19. Table 22.txt / 11,986 bytes / May 25, 2003/ Microsoft Windows Word 2000 / PC.

WHAT IS CLAIMED IS:

1. A composition-of-matter comprising a crystallized glucocerebrosidase molecule, wherein said crystallized glucocerebrosidase molecule is characterized by an X-ray diffraction capacity enabling generation of a set of structure coordinates defining a 3D structure of said glucocerebrosidase molecule or a portion thereof.

2. The composition-of-matter of claim 1, wherein said set of structure coordinates defines said 3D structure to a resolution of 2.9 angstroms or higher.

3. The composition-of-matter of claim 1, wherein an amino acid sequence of said glucocerebrosidase molecule is partially glycosylated.

4. The composition-of-matter of claim 1, wherein said crystallized glucocerebrosidase molecule is characterized by unit cell dimensions of a = about 107.7 angstroms, b = about 285.2 angstroms and c = about 91.8 angstroms.

5. The composition-of-matter of claim 1, wherein said crystallized glucocerebrosidase molecule is characterized by a crystal space group of $C222_1$.

6. The composition-of-matter of claim 1, wherein said glucocerebrosidase molecule is capable of displaying normal enzymatic activity.

7. The composition-of-matter of claim 1, wherein an amino acid sequence of said glucocerebrosidase molecule is set forth in SEQ ID NO: 1.

8. The composition-of-matter of claim 1, wherein said set of structure coordinates comprises a set of structure coordinates set forth in Table 4, 5, 6, 7, 8, 9, and/or 10.

9. A method of identifying a compound capable of correcting an impaired enzymatic activity of a mutant glucocerebrosidase molecule, the method comprising:

(a) obtaining a first set of structure coordinates, said first set of structure

coordinates defining a 3D structure of a glucocerebrosidase molecule capable of displaying normal enzymatic activity or a portion thereof;

(b) computationally generating using said first set of structure coordinates a second set of structure coordinates, said second set of structure coordinates defining a predicted 3D structure of the mutant glucocerebrosidase molecule or a portion thereof; and

(c) computationally identifying, using said second set of structure coordinates, a compound capable of interacting with the mutant glucocerebrosidase molecule in such a way as to correct the impaired enzymatic activity thereof, thereby identifying the compound capable of correcting the impaired enzymatic activity of the mutant glucocerebrosidase molecule.

10. The method of claim 9, wherein step (c) is effected further using said first set of structure coordinates.

11. The method of claim 9, further comprising biochemically qualifying a capacity of the compound to correct the impaired enzymatic activity of the mutant glucocerebrosidase molecule.

12. The method of claim 9, wherein said first set of structure coordinates defines said 3D structure at a resolution of 2.9 angstroms or higher.

13. The method of claim 9, wherein an amino acid sequence of said glucocerebrosidase molecule capable of displaying normal enzymatic activity is partially glycosylated.

14. The method of claim 9, wherein an amino acid sequence of said glucocerebrosidase molecule capable of displaying normal enzymatic activity is set forth in SEQ ID NO: 1.

15. The method of claim 9, wherein an amino acid sequence of said glucocerebrosidase molecule capable of displaying normal enzymatic activity is

composed of 497 amino acid residues, and whereas said portion of said glucocerebrosidase molecule capable of displaying normal enzymatic activity comprises a set of amino acid residues of said amino acid sequence of said glucocerebrosidase molecule having normal activity having amino acid sequence coordinates selected from the group consisting of:

(i) 76, 81, 285, 312, 314, 320, 324, 325, 336, 364–378, 423, and 433;

(ii) 244–247, and 390–397;

(iii) 20, 21, 95–100, and 404–411;

(iv) 65–67, 440–447, 460–464, 468, and 469;

(v) 360–366, 443–446, 460–467, and 484–89; and/or

(vi) 33–35, 69, 71, 450–456, 474–478, and 493–497.

16. The method of claim 9, wherein an amino acid sequence of the mutant glucocerebrosidase molecule is composed of 497 amino acid residues, and whereas said portion of the mutant glucocerebrosidase molecule comprises a set of amino acid residues of said amino acid sequence of said mutant glucocerebrosidase molecule having amino acid sequence coordinates selected from the group consisting of:

(i) 76, 81, 285, 312, 314, 320, 324, 325, 336, 364–378, 423, and 433;

(ii) 244–247, and 390–397;

(iii) 20, 21, 95–100, and 404–411;

(iv) 65–67, 440–447, 460–464, 468, and 469;

(v) 360–366, 443–446, 460–467, and 484–89; and/or

(vi) 33–35, 69, 71, 450–456, 474–478, and 493–497.

17. The method of claim 9, wherein said first set of structure coordinates comprises a set of structure coordinates set forth in Table 4, 5, 6, 7, 8, 9, and/or 10.

18. The method of claim 9, wherein an amino acid sequence of the mutant glucocerebrosidase molecule is set forth in SEQ ID NO: 2, 3, 4, 5, 6, or 7.

19. The method of claim 9, wherein said second set of structure coordinates comprises a set of structure coordinates set forth in Table 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, and/or 22.

20. The method of claim 9, wherein said glucocerebrosidase molecule capable of displaying normal enzymatic activity is a crystallized glucocerebrosidase molecule.

21. The method of claim 20, wherein said crystallized glucocerebrosidase molecule is characterized by unit cell dimensions of a = about 107.7 angstroms, b = about 285.2 angstroms and c = about 91.8 angstroms.

22. The method of claim 20, wherein said crystallized glucocerebrosidase molecule is characterized by a crystal space group of $C222_1$.

23. A computing platform capable of generating a model representing a 3D structure of a glucocerebrosidase molecule or a portion thereof, the computing platform comprising:

(a) a data-storage device storing data comprising a set of structure coordinates defining the 3D structure of the glucocerebrosidase molecule or the portion thereof; and

(b) a processing unit being for generating the model representing the 3D structure from said data stored in said data-storage device.

24. The computing platform of claim 23, wherein said set of structure coordinates defines the 3D structure at a resolution of 2.9 angstroms or higher.

25. The computing platform of claim 23, wherein an amino acid sequence of the glucocerebrosidase molecule is partially glycosylated.

26. The computing platform of claim 23, wherein the glucocerebrosidase molecule is a glucocerebrosidase molecule capable of displaying normal enzymatic activity, or is a mutant glucocerebrosidase molecule.

27. The computing platform of claim 23, wherein an amino acid sequence of the glucocerebrosidase molecule is set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, or 7.

28. The computing platform of claim 23, wherein an amino acid sequence of the glucocerebrosidase molecule is composed of 497 amino acid residues, and whereas the portion of the glucocerebrosidase molecule comprises a set of amino acid residues of said amino acid sequence having amino acid sequence coordinates selected from the group consisting of:

(i) 76, 81, 285, 312, 314, 320, 324, 325, 336, 364–378, 423, and 433;

(ii) 244–247, and 390–397;

(iii) 20, 21, 95–100, and 404–411;

(iv) 65–67, 440–447, 460–464, 468, and 469;

(v) 360–366, 443–446, 460–467, and 484–89; and/or

(vi) 33–35, 69, 71, 450–456, 474–478, and 493–497.

29. The computing platform of claim 23, wherein said set of structure coordinates comprises a set of structure coordinates set forth in Table 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, and/or 22.

30. The computing platform of claim 23, wherein the glucocerebrosidase molecule is a crystallized glucocerebrosidase molecule.

31. The computing platform of claim 30, wherein said crystallized glucocerebrosidase molecule is characterized by unit cell dimensions of a = about 107.7 angstroms, b = about 285.2 angstroms and c = about 91.8 angstroms.

32. The computing platform of claim 30, wherein said crystallized glucocerebrosidase molecule is characterized by a crystal space group of $C222_1$.

33. A method of crystallizing a glucocerebrosidase molecule, the method comprising:

(a) partially deglycosylating the glucocerebrosidase molecule, thereby generating a partially glycosylated glucocerebrosidase molecule; and

(b) subjecting said partially glycosylated glucocerebrosidase molecule to crystallization-inducing conditions, thereby crystallizing the glucocerebrosidase molecule.

34. The method of claim 33, wherein step (a) is effected by treating the glucocerebrosidase molecule with N-glycosidase F.

35. The method of claim 33, wherein an amino acid sequence of the glucocerebrosidase molecule comprises a first N-linked glycosylation consensus sequence, wherein said first N-linked glycosylation consensus sequence is attached to a sugar moiety comprising a monosaccharide or a disaccharide directly attached to said first N-linked glycosylation consensus sequence, and whereas step (a) is effected so as to leave said monosaccharide or said disaccharide attached to said first N-linked glycosylation consensus sequence.

36. The method of claim 35, wherein said sugar moiety is composed of N-acetylglucosamine moieties.

37. The method of claim 35, wherein step (a) is further effected so as to fully deglycosylate all glycosylated N-linked glycosylation consensus sequences of said amino acid sequence of said glucocerebrosidase molecule other than said first N-linked glycosylation consensus sequence of said amino acid sequence of said glucocerebrosidase molecule.

38. The method of claim 33, wherein said crystallization-inducing conditions comprise inducing evaporation of a crystallization solution containing said at least partially deglycosylated glucocerebrosidase molecule at a concentration of about 5 mg/ml, and a component selected from the group consisting of a buffer, a sodium salt, an ammonium salt, a sulfate salt, a chaotropic compound, a potassium salt, and a chloride ion.

39. The method of claim 38, wherein said buffer is a Zwitterionic buffer or an acetate buffer.

40. The method of claim 38, wherein said buffer is 2-morpholinoethanesulfonic acid buffer or sodium acetate buffer.

41. The method of claim 38, wherein said crystallization solution contains said buffer at a concentration of about 0.5 millimolar or about 0.05 molar.

42. The method of claim 38, wherein said solution of a buffer has a pH of about 6.6 or about 4.6.

43. The method of claim 38, wherein said sodium salt is sodium chloride.

44. The method of claim 38, wherein said crystallization solution contains said sodium salt at a concentration of about 0.05 molar.

45. The method of claim 38, wherein said ammonium salt is ammonium sulfate.

46. The method of claim 38, wherein said crystallization solution contains said ammonium salt at a concentration of about 0.5 molar.

47. The method of claim 38, wherein said crystallization solution contains said sulfate salt at a concentration of about 0.5 molar.

48. The method of claim 38, wherein said chaotropic compound is guanidine hydrochloride.

49. The method of claim 38, wherein said crystallization solution contains said chaotropic compound at a concentration of about 0.085 molar.

50. The method of claim 38, wherein said potassium salt is potassium chloride.

51. The method of claim 38, wherein said crystallization solution contains said potassium salt at a concentration of about 0.01 molar.

52. The method of claim 38, wherein said crystallization solution contains

said chloride ion at a concentration of about 0.06 molar.

53. The method of claim 38, wherein said crystallization solution has a pH of about 4.6.

54. The method of claim 38, wherein said inducing evaporation of said crystallization solution is effected at a temperature of about 22 degrees centigrade.

55. A computer-readable medium comprising, in a retrievable format, data including a set of structure coordinates defining a 3D structure of a glucocerebrosidase molecule or a portion thereof, wherein said set of structure coordinates defines said 3D structure at a resolution of 2.9 angstroms or higher, and/or wherein an amino acid sequence of said glucocerebrosidase molecule is partially glycosylated.

56. A computer generated model representing a 3D structure of a glucocerebrosidase molecule or a portion thereof, wherein the model represents said glucocerebrosidase molecule or a portion thereof at a resolution of 2.9 angstroms or higher, and/or wherein an amino acid sequence of said glucocerebrosidase molecule is partially glycosylated.

Gal Ehrlich
Patent Attorney
G. E. Ehrlich (1995) Ltd.
28 Bezalel Street
52521 Ramat-Gan

SEQUENCE LISTING

<110> Futerman, Anthony
Sussman, Joel
Silman, Israel
Harel, Michal
Dvir, Hay
Toker, Lilly

<120> METHODS OF IDENTIFYING GAUCHER DISEASE DRUGS

<130> 26058

<150> US 60/463,049
<151> 2003-04-16

<160> 15

<170> PatentIn version 3.2

<210> 1
<211> 497
<212> PRT
<213> Homo sapiens

<400> 1

Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1 5 10 15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
20 25 30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
35 40 45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
50 55 60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65 70 75 80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
85 90 95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
100 105 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
115 120 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
130 135 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145 150 155 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
165 170 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
180 185 190

```
Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
        195                 200                 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
    210                 215                 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225                 230                 235                 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
                245                 250                 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
            260                 265                 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
        275                 280                 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
    290                 295                 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305                 310                 315                 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
                325                 330                 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
            340                 345                 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
        355                 360                 365

Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
    370                 375                 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe Val Asp Ser
385                 390                 395                 400

Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys Gln Pro Met
                405                 410                 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
            420                 425                 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp Ala Val Ala
        435                 440                 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Arg Ser
    450                 455                 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465                 470                 475                 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp His Arg
                485                 490                 495
```

Gln

<210> 2
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (370)..(370)
<223> Asn to Ser mutant

<400> 2

```
Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1               5                  10                  15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
            20                  25                  30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
        35                  40                  45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
    50                  55                  60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65                  70                  75                  80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
                85                  90                  95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
            100                 105                 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
        115                 120                 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
    130                 135                 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145                 150                 155                 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
                165                 170                 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
            180                 185                 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
        195                 200                 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
    210                 215                 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
```

225 230 235 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
245 250 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
260 265 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
275 280 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
290 295 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305 310 315 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
325 330 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
340 345 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
355 360 365

Thr Ser Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
370 375 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe Val Asp Ser
385 390 395 400

Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys Gln Pro Met
405 410 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
420 425 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp Ala Val Ala
435 440 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Arg Ser
450 455 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465 470 475 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp His Arg
485 490 495

Gln

<210> 3
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (394)..(394)
<223> Val to Leu mutant

<400> 3

```
Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1               5                  10                  15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
            20                  25                  30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
        35                  40                  45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
    50                  55                  60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65                  70                  75                  80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
                85                  90                  95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
            100                 105                 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
        115                 120                 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
    130                 135                 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145                 150                 155                 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
                165                 170                 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
            180                 185                 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
        195                 200                 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
    210                 215                 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225                 230                 235                 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
                245                 250                 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
            260                 265                 270
```

```
His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
        275                 280                 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
    290                 295                 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305                 310                 315                 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
                325                 330                 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
            340                 345                 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
        355                 360                 365

Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
    370                 375                 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Leu Arg Asn Phe Val Asp Ser
385                 390                 395                 400

Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys Gln Pro Met
                405                 410                 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
            420                 425                 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp Ala Val Ala
        435                 440                 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Arg Ser
    450                 455                 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465                 470                 475                 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp His Arg
                485                 490                 495

Gln
```

<210> 4
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (409)..(409)
<223> Asp to His mutant

<400> 4

```
Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1               5                  10                  15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
            20                  25                  30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
        35                  40                  45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
    50                  55                  60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65                  70                  75                  80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
                85                  90                  95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
            100                 105                 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
        115                 120                 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
    130                 135                 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145                 150                 155                 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
                165                 170                 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
            180                 185                 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
        195                 200                 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
    210                 215                 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225                 230                 235                 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
                245                 250                 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
            260                 265                 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
        275                 280                 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
    290                 295                 300
```

```
Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305                 310                 315                 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
                325                 330                 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
            340                 345                 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
        355                 360                 365

Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
    370                 375                 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe Val Asp Ser
385                 390                 395                 400

Pro Ile Ile Val Asp Ile Thr Lys His Thr Phe Tyr Lys Gln Pro Met
                405                 410                 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
            420                 425                 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp Ala Val Ala
        435                 440                 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Arg Ser
    450                 455                 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465                 470                 475                 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp His Arg
                485                 490                 495

Gln
```

<210> 5
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (444)..(444)
<223> Leu to Pro mutant

<400> 5

```
Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1               5                   10                  15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
            20                  25                  30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
```

```
        35                  40                  45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
    50                  55                  60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65                  70                  75                  80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
                85                  90                  95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
            100                 105                 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
        115                 120                 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
    130                 135                 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145                 150                 155                 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
                165                 170                 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
            180                 185                 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
        195                 200                 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
    210                 215                 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225                 230                 235                 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
                245                 250                 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
            260                 265                 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
        275                 280                 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
    290                 295                 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305                 310                 315                 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
                325                 330                 335
```

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
340 345 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
355 360 365

Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
370 375 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe Val Asp Ser
385 390 395 400

Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys Gln Pro Met
405 410 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
420 425 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Pro Asp Ala Val Ala
435 440 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Arg Ser
450 455 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465 470 475 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp His Arg
485 490 495

Gln

<210> 6
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (463)..(463)
<223> Arg to Cys mutant

<400> 6

Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1 5 10 15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
20 25 30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
35 40 45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
50 55 60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65 70 75 80

```
Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
                85                  90                  95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
            100                 105                 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
        115                 120                 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
    130                 135                 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145                 150                 155                 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
                165                 170                 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
            180                 185                 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
        195                 200                 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
    210                 215                 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225                 230                 235                 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
                245                 250                 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
            260                 265                 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
        275                 280                 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
    290                 295                 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305                 310                 315                 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
                325                 330                 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
            340                 345                 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
        355                 360                 365

Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
```

370 375 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe Val Asp Ser
385 390 395 400

Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys Gln Pro Met
405 410 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
420 425 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp Ala Val Ala
435 440 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Cys Ser
450 455 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465 470 475 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp His Arg
485 490 495

Gln

<210> 7
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (496)..(496)
<223> Arg to His mutant

<400> 7

Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1 5 10 15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
20 25 30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
35 40 45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
50 55 60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65 70 75 80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
85 90 95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
100 105 110

```
Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
        115                 120                 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
    130                 135                 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145                 150                 155                 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
                165                 170                 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
            180                 185                 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
        195                 200                 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
    210                 215                 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225                 230                 235                 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
                245                 250                 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
            260                 265                 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
        275                 280                 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
    290                 295                 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305                 310                 315                 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
                325                 330                 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
            340                 345                 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
        355                 360                 365

Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
    370                 375                 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe Val Asp Ser
385                 390                 395                 400

Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys Gln Pro Met
                405                 410                 415
```

```
Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
            420                 425                 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp Ala Val Ala
        435                 440                 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Arg Ser
    450                 455                 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465                 470                 475                 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp His Arg
                485                 490                 495

Gln
```

<210> 8
<211> 497
<212> PRT
<213> Homo sapiens

<400> 8

```
Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1               5                   10                  15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
            20                  25                  30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
        35                  40                  45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
    50                  55                  60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65                  70                  75                  80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
                85                  90                  95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
            100                 105                 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
        115                 120                 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
    130                 135                 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145                 150                 155                 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
                165                 170                 175
```

```
Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
            180                 185                 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
        195                 200                 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
    210                 215                 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225                 230                 235                 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
                245                 250                 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
            260                 265                 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
        275                 280                 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
    290                 295                 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305                 310                 315                 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
                325                 330                 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
            340                 345                 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
        355                 360                 365

Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
    370                 375                 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe Val Asp Ser
385                 390                 395                 400

Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys Gln Pro Met
                405                 410                 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
            420                 425                 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp Ala Val Ala
        435                 440                 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Arg Ser
    450                 455                 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
```

465 470 475 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp Arg Arg
485 490 495

Gln

<210> 9
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (370)..(370)
<223> Asn ot Ser mutant

<400> 9

Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1 5 10 15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
20 25 30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
35 40 45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
50 55 60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65 70 75 80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
85 90 95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
100 105 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
115 120 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
130 135 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145 150 155 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
165 170 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
180 185 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
195 200 205

```
Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
    210                 215                 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225                 230                 235                 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
                245                 250                 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
            260                 265                 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
        275                 280                 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
    290                 295                 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305                 310                 315                 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
                325                 330                 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
            340                 345                 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
        355                 360                 365

Thr Ser Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
    370                 375                 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe Val Asp Ser
385                 390                 395                 400

Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys Gln Pro Met
                405                 410                 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
            420                 425                 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp Ala Val Ala
        435                 440                 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Arg Ser
    450                 455                 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465                 470                 475                 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp Arg Arg
                485                 490                 495

Gln
```

<210> 10
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (394)..(394)
<223> Val to Leu mutant

<400> 10

Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1 5 10 15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
20 25 30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
35 40 45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
50 55 60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65 70 75 80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
85 90 95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
100 105 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
115 120 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
130 135 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145 150 155 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
165 170 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
180 185 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
195 200 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
210 215 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225 230 235 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
245 250 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
260 265 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
275 280 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
290 295 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305 310 315 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
325 330 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
340 345 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
355 360 365

Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
370 375 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Leu Arg Asn Phe Val Asp Ser
385 390 395 400

Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys Gln Pro Met
405 410 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
420 425 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp Ala Val Ala
435 440 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Arg Ser
450 455 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465 470 475 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp Arg Arg
485 490 495

Gln

<210> 11
<211> 497
<212> PRT
<213> Homo sapiens

<220>

<221> misc_feature
<222> (409)..(409)
<223> Asp to His mutant

<400> 11

```
Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1               5                  10                  15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
            20                  25                  30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
        35                  40                  45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
    50                  55                  60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65                  70                  75                  80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
                85                  90                  95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
            100                 105                 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
        115                 120                 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
    130                 135                 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145                 150                 155                 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
                165                 170                 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
            180                 185                 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
        195                 200                 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
    210                 215                 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225                 230                 235                 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
                245                 250                 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
            260                 265                 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
```

275 280 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
290 295 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305 310 315 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
325 330 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
340 345 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
355 360 365

Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
370 375 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe Val Asp Ser
385 390 395 400

Pro Ile Ile Val Asp Ile Thr Lys His Thr Phe Tyr Lys Gln Pro Met
405 410 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
420 425 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp Ala Val Ala
435 440 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Arg Ser
450 455 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465 470 475 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp Arg Arg
485 490 495

Gln

<210> 12
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (444)..(444)
<223> Leu to Pro

<400> 12

Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1 5 10 15

```
Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
            20                  25                  30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
        35                  40                  45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
    50                  55                  60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65                  70                  75                  80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
                85                  90                  95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
            100                 105                 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
        115                 120                 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
    130                 135                 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145                 150                 155                 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
                165                 170                 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
            180                 185                 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
        195                 200                 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
    210                 215                 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225                 230                 235                 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
                245                 250                 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
            260                 265                 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
        275                 280                 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
    290                 295                 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305                 310                 315                 320
```

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
325 330 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
340 345 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
355 360 365

Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
370 375 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe Val Asp Ser
385 390 395 400

Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys Gln Pro Met
405 410 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
420 425 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Pro Asp Ala Val Ala
435 440 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Arg Ser
450 455 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465 470 475 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp Arg Arg
485 490 495

Gln

<210> 13
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (463)..(463)
<223> Arg to Cys mutant

<400> 13

Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1 5 10 15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
20 25 30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
35 40 45

```
Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
 50                  55                  60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65                  70                  75                  80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
                85                  90                  95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
            100                 105                 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
        115                 120                 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
    130                 135                 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145                 150                 155                 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
                165                 170                 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
            180                 185                 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
        195                 200                 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
    210                 215                 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225                 230                 235                 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
                245                 250                 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
            260                 265                 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
        275                 280                 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
    290                 295                 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305                 310                 315                 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
                325                 330                 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
            340                 345                 350
```

```
Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
        355                 360                 365

Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
    370                 375                 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe Val Asp Ser
385                 390                 395                 400

Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys Gln Pro Met
                405                 410                 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
            420                 425                 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp Ala Val Ala
        435                 440                 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Cys Ser
    450                 455                 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465                 470                 475                 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp Arg Arg
                485                 490                 495

Gln
```

<210> 14
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (496)..(496)
<223> Arg to His mutant

<400> 14

```
Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser Val Val Cys
1               5                   10                  15

Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro Thr Phe Pro
            20                  25                  30

Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser Gly Arg Arg
        35                  40                  45

Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr Gly Thr Gly
    50                  55                  60

Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys Val Lys Gly
65                  70                  75                  80

Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile Leu Ala Leu
```

85 90 95

Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe Ser Glu Glu
100 105 110

Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser Cys Asp Phe
115 120 125

Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp Phe Gln Leu
130 135 140

His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys Ile Pro Leu
145 150 155 160

Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser Leu Leu Ala
165 170 175

Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly Ala Val Asn
180 185 190

Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr His Gln Thr
195 200 205

Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala Glu His Lys
210 215 220

Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser Ala Gly Leu
225 230 235 240

Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro Glu His Gln
245 250 255

Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala Asn Ser Thr
260 265 270

His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg Leu Leu Leu
275 280 285

Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala Ala Lys Tyr
290 295 300

Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu Ala Pro Ala
305 310 315 320

Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn Thr Met Leu
325 330 335

Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu Gln Ser Val
340 345 350

Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His Ser Ile Ile
355 360 365

Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp Asn Leu Ala
370 375 380

Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe Val Asp Ser
385 390 395 400

Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys Gln Pro Met
405 410 415

Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu Gly Ser Gln
420 425 430

Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp Ala Val Ala
435 440 445

Leu Met Asn Pro Asp Gly Ser Ala Val Val Val Val Leu Asn Arg Ser
450 455 460

Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val Gly Phe Leu
465 470 475 480

Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu Trp Arg Arg
485 490 495

Gln

<210> 15
<211> 3
<212> PRT
<213> Artificial sequence

<220>
<223> N-linked glycosylation consensus sequence

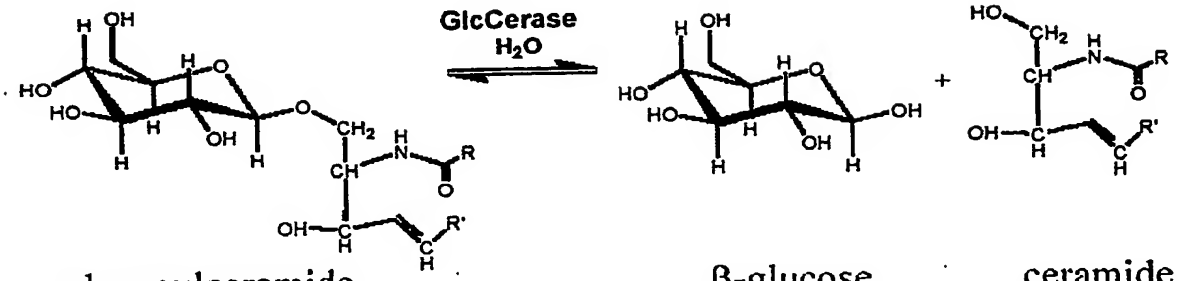
<220>
<221> misc_feature
<222> (2)..(2)
<223> Any amino acid

<220>
<221> misc_feature
<222> (3)..(3)
<223> Ser or Thr

<400> 15

Asn Xaa Xaa
1

Fig. 1a



glucosylceramide

$R,R'=(CH_2)_{13\text{-}21}CH_3$

β-glucose

ceramide

Fig. 1b

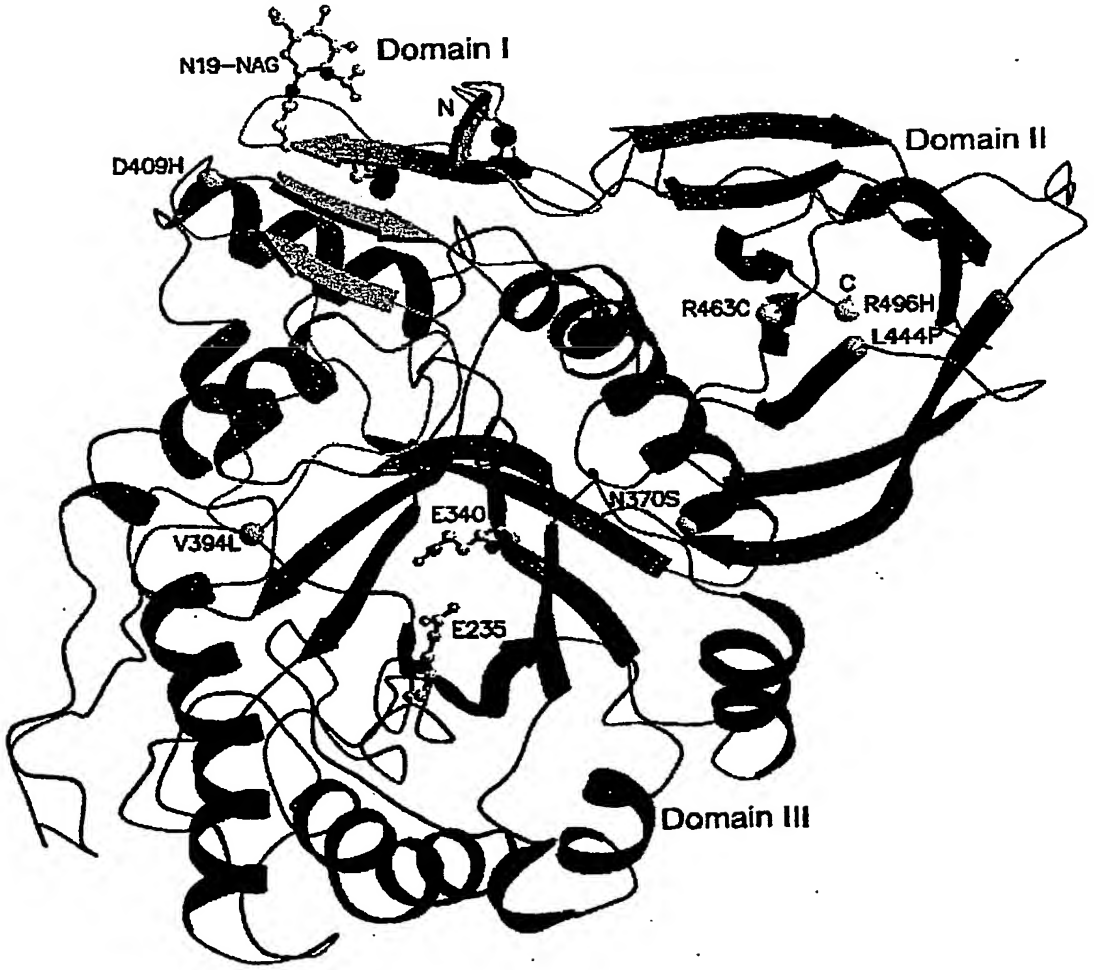


Fig. 1c

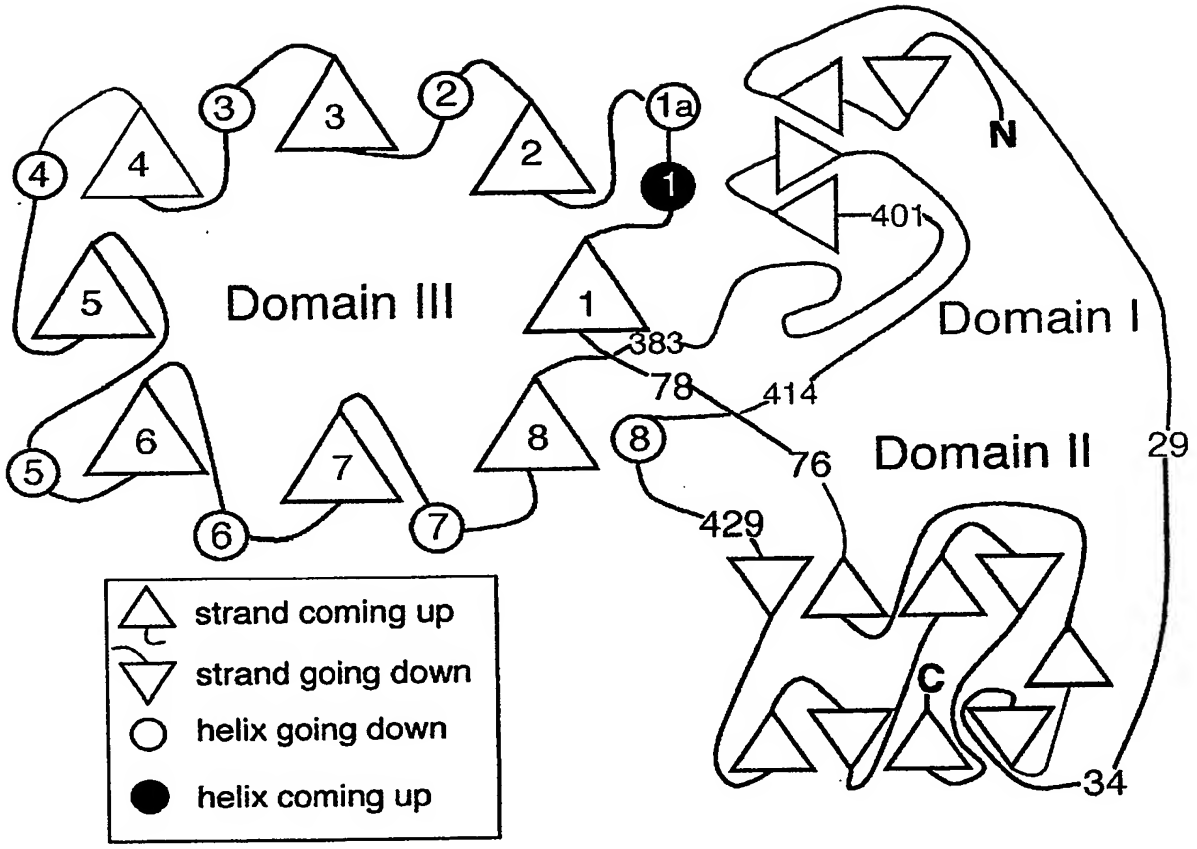

Domain III
Domain I
Domain II
N
C
1a
1
2
3
4
5
6
7
8
401
383
78
414
76
429
29
34
strand coming up
strand going down
helix going down
helix coming up

Fig. 1d

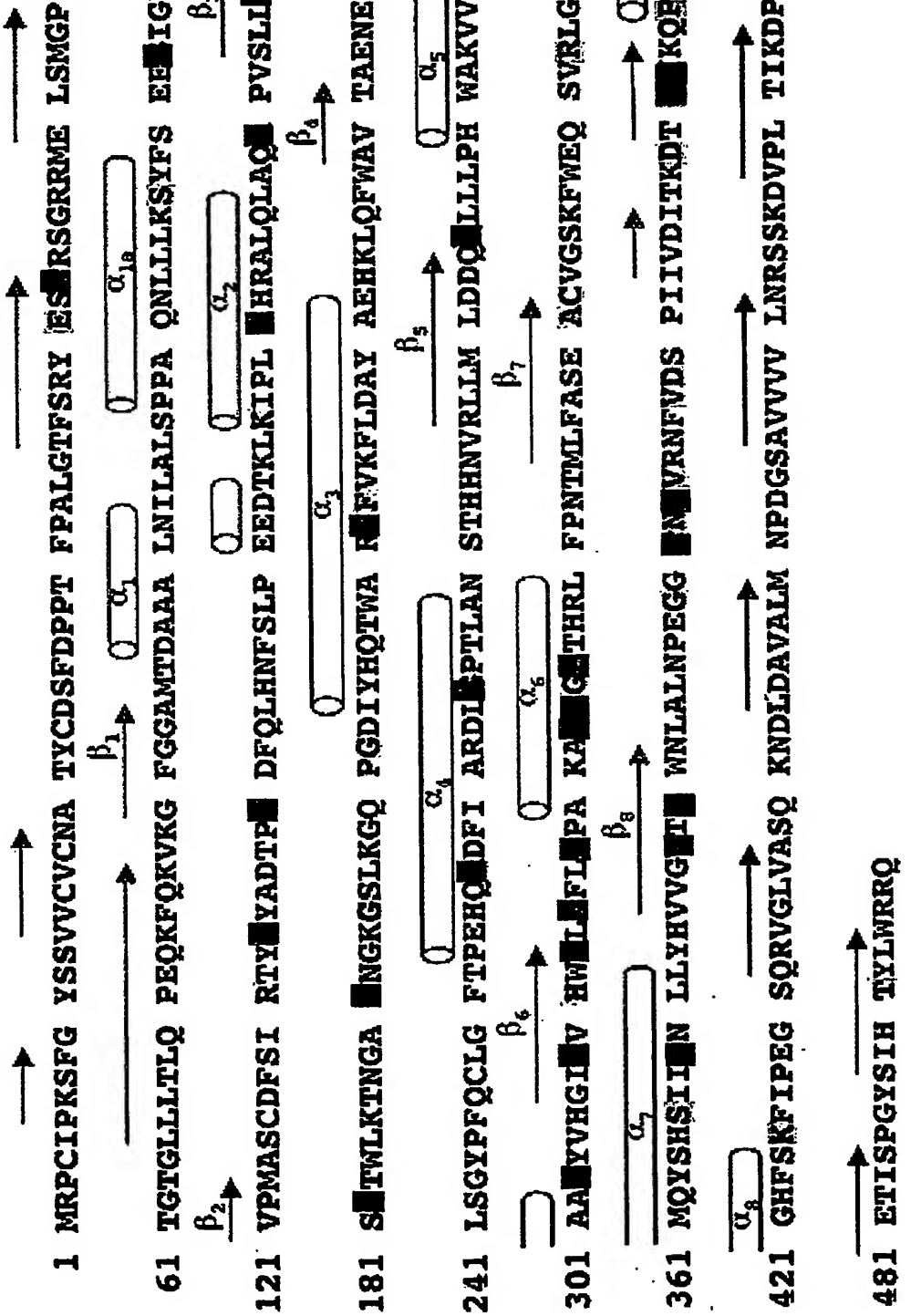

1 MRPCIPKSFG YSSVVCVCNA TYCDSFDPPT FPALGTFSRY ES RSGRRME LSMGPIQANH
61 TGTGLLLTLQ PEQKFQKVKG FGGAMTDAAA LNILALSPPA QNLLLKSYFS EE IGYNIER
121 VPMASCDFSI RTY YADTP DFQLHNFSLP EEDTKLKIPL HRALQLAQ PVSLI SPWT
181 S TWLKTNGA NGKGSLKGQ PGDIYHQTWA K FVKFLDAY AEHKLQFWAV TAENEPSAGL
241 LSGYPFQCLG FTPEHQ DFI ARDI PTLAN STHHNVRLLM LDDQ LLLPH WAKVVLTDPE
301 AA YVHGI V HW L FL PA KA G THRL FPNTMLFASE ACVGSKFWEQ SVRLGSWDRG
361 MQYSHSII N LLYHVVG T WNLALNPEGG N VRNFWDS PIIVDITKDT KQPM HL
421 GHFSKFIPEG SQRVGLVASQ KNDIDAVALM NPDGSAVVVV LNRSSKDVPL TIKDPAV FL
481 ETISPGYSIH TYLWRRQ

Fig. 1e

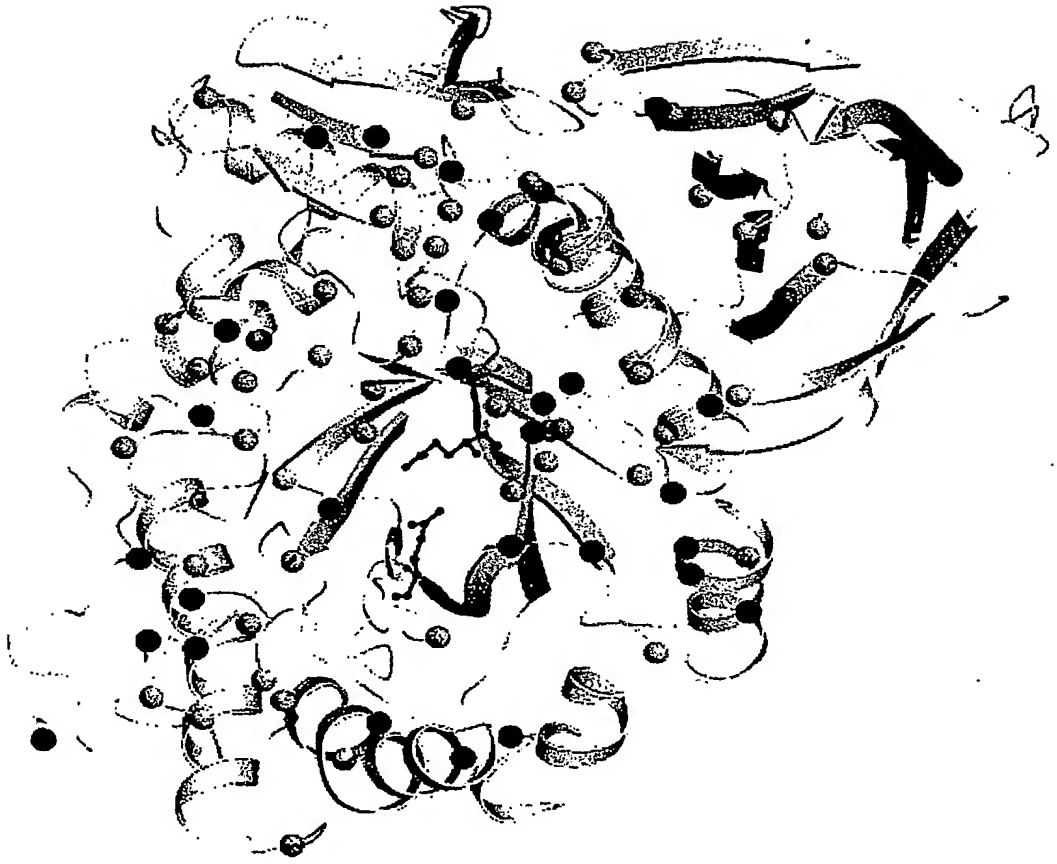


Fig. 2a

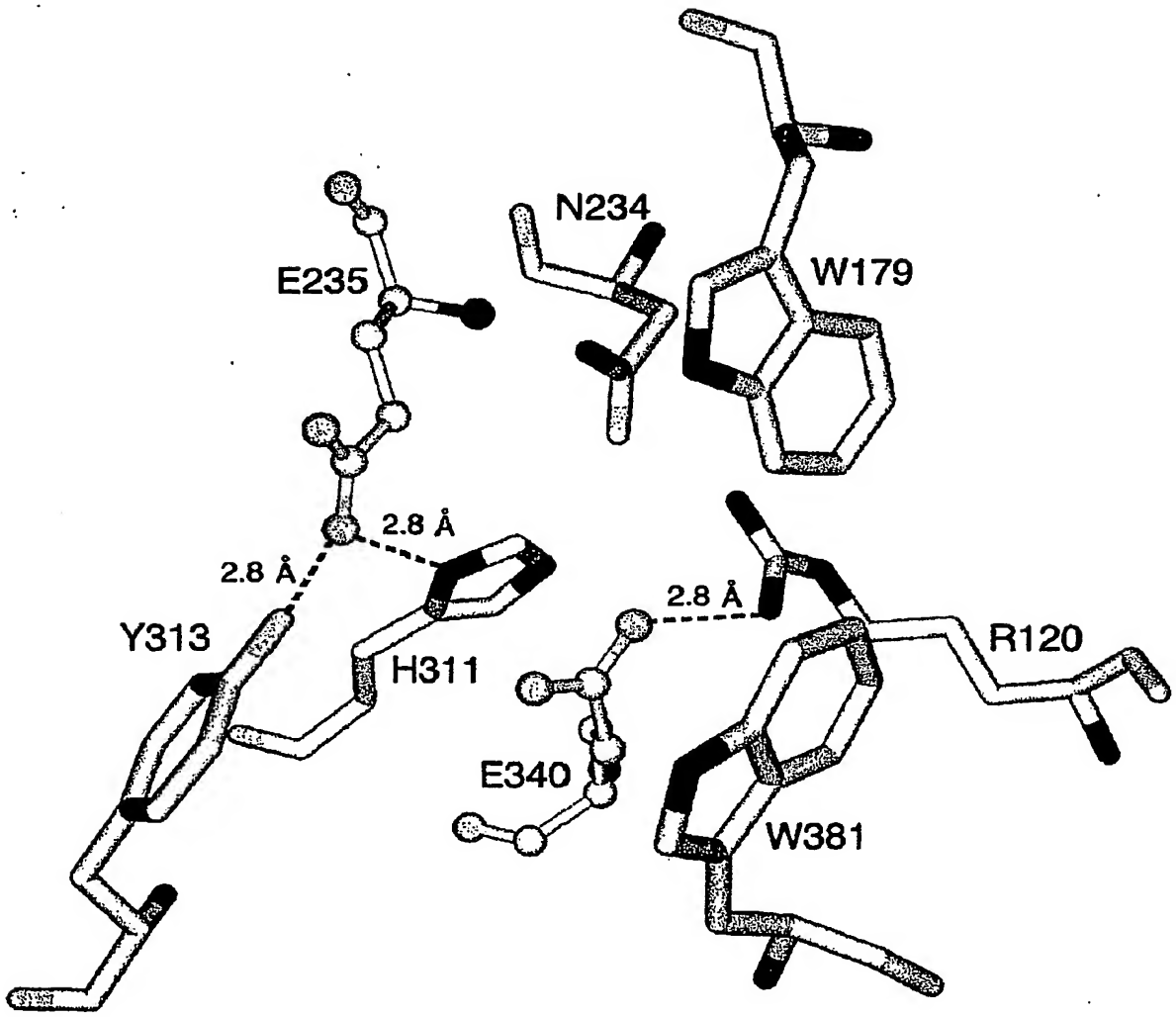
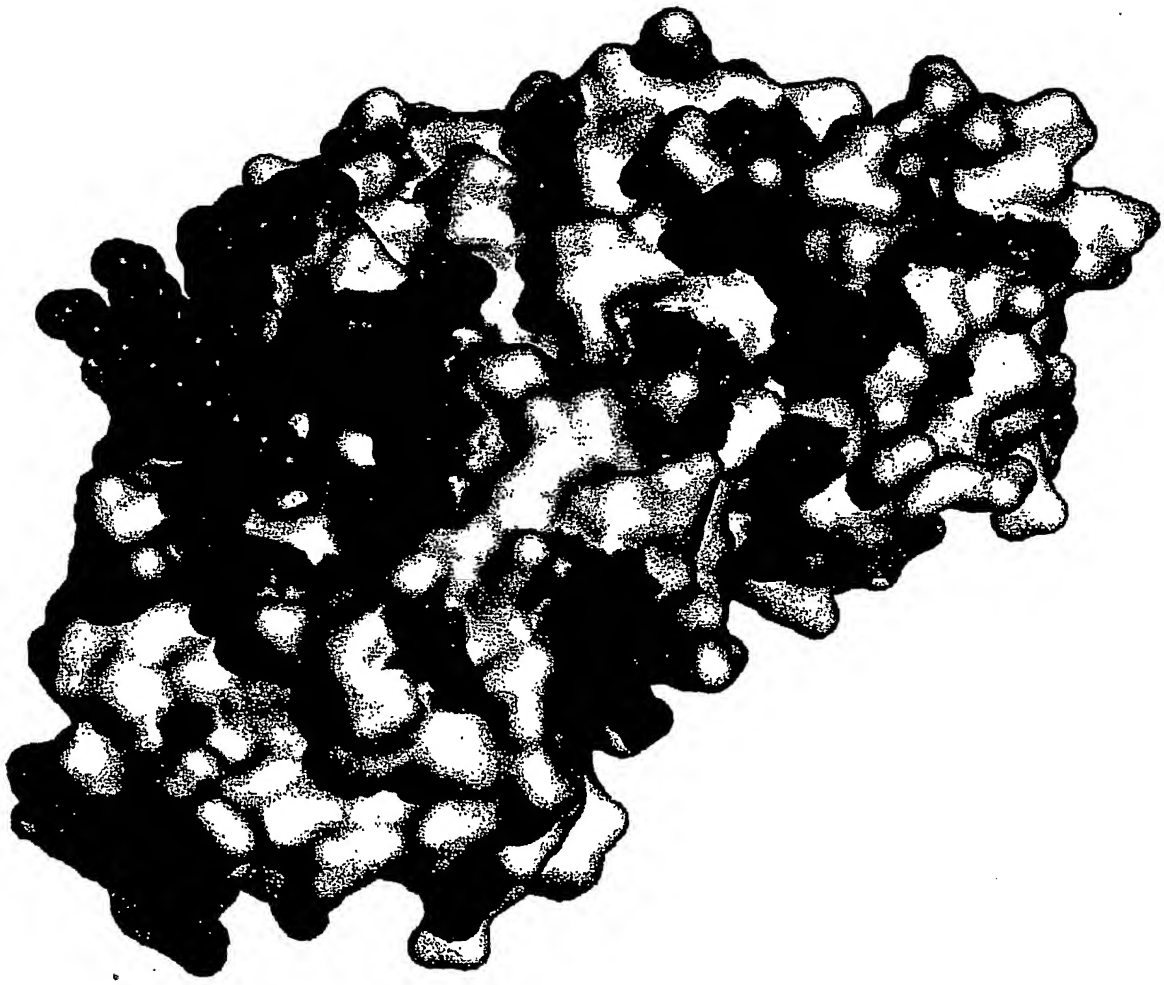


Fig. 2b



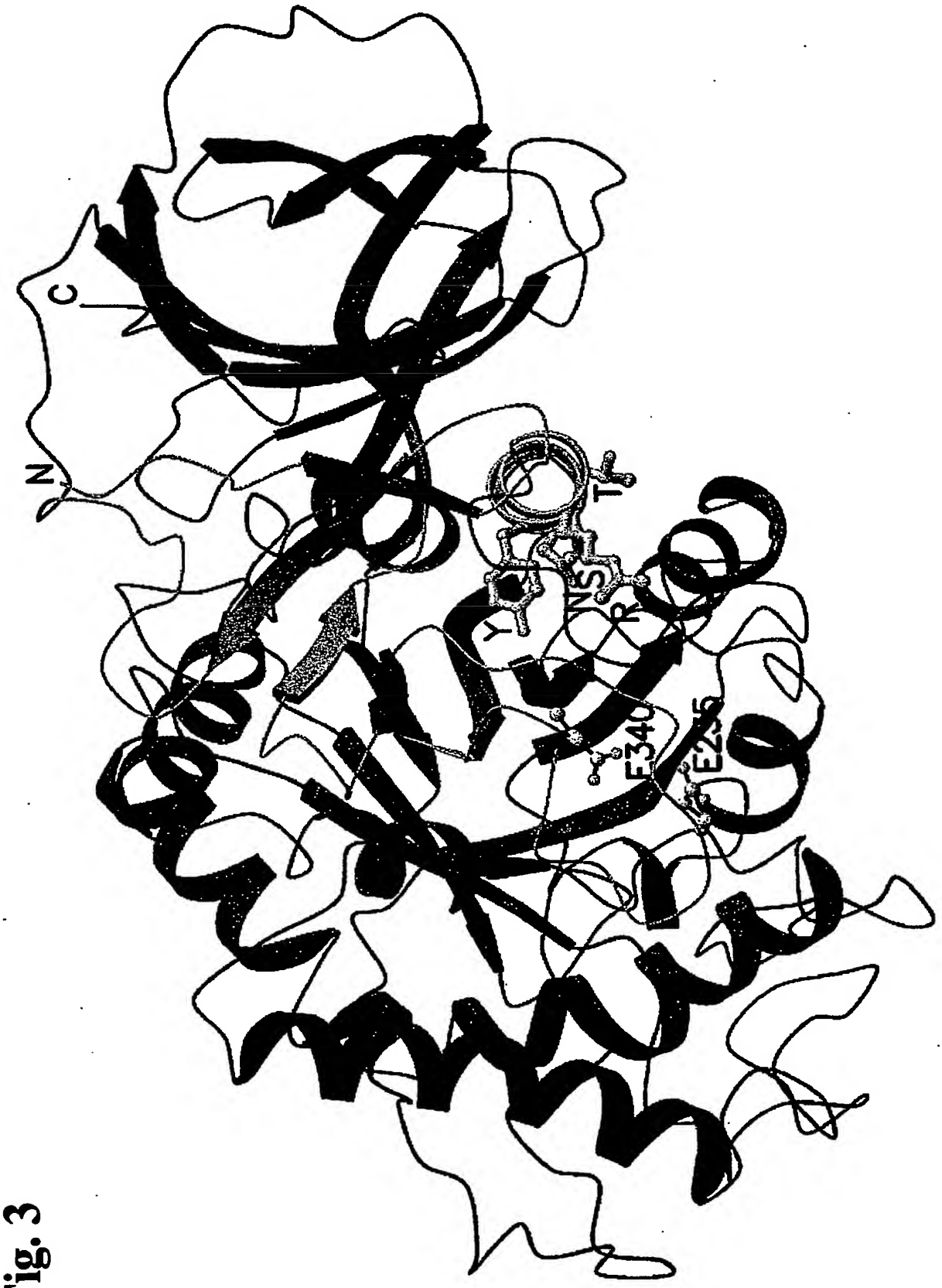


Fig. 3

Fig. 4